

From: Swope, Sheridan  
Sent: Thursday, June 09, 2005 5:56 PM  
To: STIC-Biotech/ChemLib  
Subject: 10/649,273

For 10/649,273, pls search and interference search:

SID 2: full-length against the NT data bases  
148-414 against the NT data bases  
176-414 against the NT data bases

Sheridan Swope, Ph.D.  
Patent Examiner, AU 1652  
Recombinant Enzymes  
571-272-0943 (voice)  
E02B71 Remsen Bld (Office)  
E02C70 Remsen Bld (Mailbox)

Paul Schulwitz

\*\*\*\*\*

STAFF USE ONLY

Searcher: \_\_\_\_\_  
Searcher Phone: 2- \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date Completed: 6/20 \_\_\_\_\_  
Searcher Prep/Rev. Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search

NA#: \_\_\_\_\_ AA#: \_\_\_\_\_  
Interference: \_\_\_\_\_ SPDI: \_\_\_\_\_  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure#: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other(Specify): \_\_\_\_\_

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 16, 2005, 15:19:45 ; Search time 4951.35 Seconds

(without alignments)  
4051.513 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125  
Sequence: 1 MLIRTAGVFKPKSRKRVY.....DISKEVGEASIKYQQLMEI 414

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlp  
-Q/cgml 1/USPTO.spool\_p/US10649273/runat.15062005.111416.6030/app\_query.fasta.1.1429  
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPT=0 -LOPEXT=0  
-INITs-bits -START=1 -END=1 -MATRIX=bloum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10649273 @CGML 1.1 7742 @runat.15062005.111416.6030 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:.\*  
1: gb ba:.\*  
2: gb htg:.\*  
3: gb ln:.\*  
4: gb om:.\*  
5: gb ov:.\*  
6: gb pat:.\*  
7: gb ph:.\*  
8: gb pl:.\*  
9: gb pr:.\*  
10: gb ro:.\*  
11: gb ste:.\*  
12: gb sy:.\*  
13: gb un:.\*  
14: gb vi:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID     | Description |
|------------|--------|-------------|--------|-----------|-------------|
| 1          | 2125   | 100.0       | 1908   | BC011904  | Homo sapi   |
| 2          | 2125   | 100.0       | 2197   | AR428803  | Sequence    |
| 3          | 2090.5 | 98.4        | 1387   | AR428808  | Sequence    |
| 4          | 2090.5 | 98.4        | 1387   | HSA295148 | Homo sapi   |

|    |        |      |        |    |            |                    |
|----|--------|------|--------|----|------------|--------------------|
| 5  | 2088   | 98.3 | 1245   | 6  | AX664697   | Sequence           |
| 6  | 2088   | 98.3 | 1820   | 6  | AX664695   | Sequence           |
| 7  | 1944   | 91.5 | 2208   | 6  | AX713716   | Sequence           |
| 8  | 1944   | 91.5 | 2208   | 6  | AK055441   | Homo sapi          |
| 9  | 1835   | 86.4 | 1844   | 10 | BC058172   | Mus muscu          |
| 10 | 1747   | 82.2 | 1416   | 6  | ARS41929   | Sequence           |
| 11 | 1725   | 81.2 | 1546   | 10 | BC078974   | Sequence           |
| 12 | 1385   | 65.2 | 1526   | 6  | AR428809   | Sequence           |
| 13 | 1362   | 64.1 | 1522   | 5  | BX934991   | Gallus ga          |
| 14 | 1273   | 59.9 | 1017   | 10 | BC038910   | Mus muscu          |
| 15 | 1208   | 56.8 | 1558   | 5  | BX930963   | Gallus ga          |
| 16 | 1193.5 | 56.2 | 84115  | 9  | AC013468   | Homo sapi          |
| 17 | 1186.5 | 55.8 | 14364  | 6  | AR428807   | Sequence           |
| 18 | 995.5  | 46.8 | 249601 | 2  | AC114153   | Rattus no          |
| 19 | 995.5  | 46.8 | 308652 | 2  | AC121478   | Rattus no          |
| 20 | 970    | 45.6 | 256751 | 10 | AC122925   | Mus muscu          |
| 21 | 950.5  | 44.7 | 1109   | 10 | BC051211   | Mus muscu          |
| 22 | 938    | 44.1 | 860    | 5  | BX930694   | Gallus ga          |
| 23 | 827    | 38.9 | 875    | 6  | C0721898   | Sequence           |
| 24 | 751.5  | 35.4 | 121251 | 5  | AL591593   | Zebrafish          |
| 25 | 714.5  | 33.6 | 1576   | 3  | AY051882   | Drosophila         |
| 26 | 714.5  | 33.6 | 1601   | 6  | C0606432   | Sequence           |
| 27 | 710    | 33.4 | 1474   | 3  | AK113378   | Clona int          |
| 28 | 709    | 33.4 | 117322 | 5  | AL672217   | Zebrafish          |
| 29 | 700.5  | 33.0 | 1385   | 6  | BD157102   | Primer fo          |
| 30 | 700.5  | 33.0 | 1385   | 6  | AX878239   | Sequence           |
| 31 | 700.5  | 33.0 | 1385   | 6  | AK027836   | Homo sapi          |
| 32 | 687.5  | 32.4 | 3656   | 6  | C0606431   | Sequence           |
| 33 | 687.5  | 32.4 | 14679  | 2  | AC018262   | Drosophila         |
| 34 | 687.5  | 32.4 | 180263 | 3  | AC010671   | Drosophila         |
| 35 | 687.5  | 32.4 | 207432 | 3  | AE003513   | Drosophila         |
| 36 | 668.5  | 31.5 | 1443   | 8  | AY024338   | Arabidops          |
| 37 | 668.5  | 31.5 | 1474   | 8  | AY117283   | Arabidops          |
| 38 | 668.5  | 31.5 | 1567   | 8  | AY063864   | Arabidops          |
| 39 | 662.5  | 31.2 | 1557   | 8  | AY084577   | Arabidops          |
| 40 | 656    | 30.9 | 1672   | 8  | AK070912   | Oryza sat          |
| 41 | 629    | 29.6 | 571    | 6  | BD154975   | Primer fo          |
| 42 | 629    | 29.6 | 571    | 6  | AX874913   | Sequence           |
| 43 | 618    | 29.1 | 110000 | 2  | AP06501_08 | Continuation (9 of |
| 44 | 577.5  | 27.2 | 260271 | 1  | AE017258   | Wolbachia          |
| 45 | 568.5  | 26.8 | 333800 | 1  | SME591792  | Sinorhizo          |

## ALIGNMENTS

|            |  |         |      |                 |
|------------|--|---------|------|-----------------|
| RESULT 1   | BC011904   | 1908 bp | mRNA | PRI 23-DEC-2003 |
| LOCUS      | BC011904   |         |      |                 |
| DEFINITION | Homo sapiens O-6-allylglucoprotein endopeptidase-like 1, mRNA (CDNA  |         |      |                 |
| ACCESSION  | clone MGC:20293 IMAGE:4121450), complete cds.                        |         |      |                 |
| VERSION    | BC011904.2 GI:40225818   |         |      |                 |
| KEYWORDS   | MGC.   |         |      |                 |
| SOURCE     | Homo sapiens (human)   |         |      |                 |
| ORGANISM   | Homo sapiens   |         |      |                 |
| REFERENCE  | Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;    |         |      |                 |
| AUTHORS    | Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.          |         |      |                 |
|            | 1 (bases 1 to 1908)  |         |      |                 |
|            | Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,             |         |      |                 |
|            | Klausner,R.D., Collins,P.S., Wagner,L., Shemen,C.M., Schuler,G.D.,   |         |      |                 |
|            | Altschul,S.F., Zeeberg,B., Bucrow,K.H., Scheffer,C.F., Bat,N.K.,     |         |      |                 |
|            | Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,P.,      |         |      |                 |
|            | Diachenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,         |         |      |                 |
|            | Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,              |         |      |                 |
|            | Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,             |         |      |                 |
|            | Carrinci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,     |         |      |                 |
|            | Abramson,R.D., Muliahy,S.J., Bosak,S.A., McEwan,P.J.,                |         |      |                 |
|            | McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richard,S.,               |         |      |                 |
|            | Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Nulya,S.W.,             |         |      |                 |
|            | Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,        |         |      |                 |
|            | Fahy,J., Helton,E., Kettman,M., Madan,A., Young,A.C., Shcherenko,Y., |         |      |                 |
|            | Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shcherenko,Y.,         |         |      |                 |
|            | Boulford,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D.,             |         |      |                 |

**TITLE**  
Dickson, M.C., Rodriguez, A.C., Gimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.E., Scherich, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

**JOURNAL**  
PUBLISHED  
12477932  
2 (bases 1 to 1908)  
Strausberg, R.  
Direct Submission  
Submitted (30-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

**REMARK**  
COMMENT  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
On Dec 19, 2003 this sequence version replaced gi:15080281.  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@hgrl.nih.gov](mailto:nisc_mgc@hgrl.nih.gov)  
Akher, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, J., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R., Maduro, O.L., Masello, C., Maekari, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stankis, S., Thomas, P.J., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 28 Row: 1 Column: 22.  
Location/Qualifiers

**FEATURES**  
source

1..1908  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:20293 IMAGE:4121450"  
/tissue\_type="Muscle, rhabdomyosarcoma"  
/clone\_lib="NIH-MGC\_17"  
/lab\_host="DH10B-R"  
/note="Vector: pOT87"  
1..1908  
/gene="OSGEP1"  
/db\_xref="LocusID:64172"  
104..1348  
/gene="OSGEP1"  
/product="OSGEP1 protein"  
/protein\_id="AAH11904.1"  
/db\_xref="GI:15080282"  
/db\_xref="LocusID:64172"  
/translation="MLLTAGVFPKSRKRYEFLRSFNFGTLFLKIVGLIET  
SCDPTAAVDERNGVGEAHSOTVEALITGGIVPPAOLREHRIORIVQELSLSS  
GSPSLSAIATTIKPGLALSLGVLSFSLQVGLKPPPIPIHMEAAHLLTRLNK  
VERPPIVLISGCHLALVQVSDPRLILKSLDIAGDMLDVYARLSLTKRPGCT  
WSGKAIEHAKQNRFPDIKPLPHAKQDSFTGLQVTDKTIKKKEGEGIEKG  
QILSSADIAATVQHTMAHLVKRTHRAILFCQRDLIPNNMVLVASGVASNFYLR  
RALEILTNAOCTILCPPLCTDNGIMIMANGIERIRLAGILHIDIEGIRYBKPL  
GVDISKEVGEASIKVPLKMEI"  
215..1111  
/gene="OSGEP1"  
/note="Peptidase M22; Region: Glycoprotease family"  
/db\_xref="CD:pfam0814"

**ORIGIN**  
Alignment Scores:

Pred. No.: 6,17e-171 Length: 1908  
Score: 2125.00 Matches: 414  
Percent Similarity: 100.00% Conserved: 0  
Best local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0

US-10-649-273-2 (1-414) x BC011904 (1-1908)

QY 1 MetleuileuThrlvsthrAlaGlyValPhepheyPserSerlyArgValTyr 20  
DB 104 AAGCTAAATCTTGACTTAAGCTGCGAGAGTTTTTTTAAACATCAAGAAAGACTTAT 163  
QY 21 GlupheleuArgSerPheasnPhehisProGlyThrleupheleuHislyeIleValleu 40  
DB 164 GAATTTTAAAGACTTTTAAATTTTCACTCGAAGACATTTCTTCAATTAATAGTATG 223  
QY 41 GlylleGluThrSerCyaspasPthrAlaAlaValaIaapGluThrGlyAsnVal 60  
DB 224 GGAATTAAGAACTAGTTGATGATACAGACAGCTCTGCTGTGATGAACCTGAAATGTG 283  
QY 61 leuGlyGluAlaIleHisSerGlnThrGluValHisleuysThrGlyGlyIleValPro 80  
DB 284 TTGAGAGAGCAATACATCTCCAACTGAAGTTATTTAAACAGCGTGAATGTCTCT 343  
QY 81 ProAlaIaGlnGlnleuHisArgGluAsnIleGlnArgIleValGlnAlaIeuser 100  
DB 344 CCAGACAGCTCAACAGCTTCCACAGAAATTTTCAAGAAATGACAAAGAACTTTCT 403  
QY 101 AlaSerGlyValSerProSerAspSerleuSerAlaIlealThrThrIleysProGlyleu 120  
DB 404 GCCAGTGAGCTCTCCAAAGTACCTCTCAGCAATTCATACATCAATCAACAGACTT 463  
QY 121 AlaIeuserleuGlyValGlyleuSerPheSerleuGlnleuValGlyleuIleuys 140  
DB 464 GCTTAAAGCCCTGGAGGAGGCTTATTAAGCTTAAAGCTGAGGACAGTAAAG 523  
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaIeuthrIleArgleuThrAsnlys 160  
DB 524 CCATTCATTCCTCATTCATCATATGAGGCTCATGACTTATTAAGTTACCAATAAA 583  
QY 161 ValGluPheProPheleuValleuIleuserGlyGlyHisCySleuIleuAlaIeul 180  
DB 584 GTAAATTTCTTTTATTAAGTTCTTTTATTAAGTTCTGAGGTCATCTGTGGCATTA 643  
QY 181 GlnGlyValSerAspPheleuIleuGlyIleuserleuAspIleAlaProGlyAspMet 200  
DB 644 CAAGAGTTTCAGATTTTCTGCTTCTTGAAAGCTTTGACATAGCACAGTCAATG 703  
QY 201 leuAspIleValAlaArgArgleuSerleuIleHisIleProGluCySerThrMetSer 220  
DB 704 CTGACAAAGTGCGAGAGAGACTTTCTTTTAAATAACCTCAGAGTCTCCACATGAGT 763  
QY 221 GlyGlyIleValAlaIleGluHisIleuAlaIleGlnIleAsnArgPheHisPheAspIle 240  
DB 764 GGTGGAAAGCCATAGAAACATTTGGCCAAAGAAATATTAATTTGACATCAA 823  
QY 241 ProProleuHisIleAlaIleAsnCyaspPheSerPheThrGlyleuGlnHisValThr 260  
DB 824 CCTCCCTTGATCATGCTAAAGAAATTTCTTTTCTTCTTCTGACCTTCAACACCTTACT 883  
QY 261 AspIleIleIleMetIleGlyGlyGluGluGluGluGluGluGluGluGluGluGlu 280  
DB 884 GATAAATTAATAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 943  
QY 281 SerAlaIaAspIleAlaIa 300  
DB 944 TACAGACAGACATTTGCTGCTGACAGTACAGACACATGCAATGCACTTGTGAAAGA 1003  
QY 301 ThrHisArgAlaIaIeIleuPheCyalyGlnArgAspIleuPProGlnAsnAsnAlaVal 320  
DB 1004 ACACATGGGCTATTTCTTTTGTAAAGACAGAGACTTGTACCTCAAAATATACAGTA 1063

QY 321 LeuValAlaSerGlyValAlaSerAsnPheTyrIleArgAlaLeuGluIleLeu 340  
 DB 1064 CTGGTTCATCTGCTGCTGCGCAAGTAACTTCTATATCCGACAGCTCTGGAAATTTTA 1123  
 QY 341 ThrAsnAlaThrGlnCysThrIleuLeuCysProProAlaGluCysThrAspAsnGly 360  
 DB 1124 ACAAGCGAACACAGTCACCTTGTTGTGTCTCTCCAGACTATGACATGAATATGGC 1183  
 QY 361 IleMetIleAlaTTPAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
 DB 1184 ATTATGATTCATGAGTAATGATGATTTGAAGACTAGCTGCTGGCATTTTATCATAC 1243  
 QY 381 IleGluGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerLysGluVal 400  
 DB 1244 ATGAAGGATCCGCTATGAAACCAAAATGTCTCTTGAGATGACATATCAAAAGAGTT 1303  
 QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414  
 DB 1304 GGAGAGCTTCATAAAGTACCAATTAATAATGAGATA 1345  
 RESULT 2  
 AR428803 AR428803 2197 bp DNA linear PAT 18-DEC-2003  
 LOCUS Sequence 1 from patent US 6642041.  
 DEFINITION AR428803  
 ACCESSION AR428803  
 VERSION AR428803.1 GI:40188589  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 2197)  
 AUTHORS Chen, J., Feder, J.N., Nelson, T.C., Krystek, S.R. and Duclos, F.  
 TITLE Polynucleotides encoding a novel metalloprotease, MP-1  
 JOURNAL Patent: US 6642041-A 1 04-NOV-2003;  
 FEATURES Location/Qualifiers  
 source 1..2197  
 /organism="unknown"  
 /mol\_type="genomic DNA"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 7.37e-171 Length: 2197  
 Score: 2125.00 Matches: 414  
 Percent Similarity: 100.00% Mismatches: 0  
 Best Local Similarity: 100.00% Indels: 0  
 Query Match: 6 Gaps: 0  
 DB: US-10-649-273-2 (1-414) x AR428803 (1-2197)  
 QY 1 MetIleuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20  
 DB 231 ATGCTATCTTGAAGTAAAGTACGACGAGGTTTATTTTAAACATCAAAAAGAGTTTAT 290  
 QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40  
 DB 291 GAATTTTAAAGAACTTTTAAATTTTCAATCCCTGGAACACATATTTCTCAATAAATGATATG 350  
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60  
 DB 351 GGAATTAACCTAGTGTGATGATACAGCACTGCTGTGGTGAATGAAATGGAATGTG 410  
 QY 61 LeuGlyValAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80  
 DB 411 TTGGGAGACAGCAATACATTCCTCAAGTGAAGTTCATTTAAAAACAGGTGGATTTCTCT 470  
 QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
 DB 471 CCAAGCACTCAACAGCTTCAACAGAAATATTCACCAAAATAGTACCAAGAAAGCTCTTCT 530  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrTrpIleLysProGlyLeu 120  
 DB 531 GCCAGTGAAGTCTCTCCAGATGACCTCTCAAGCAATATGCACTACCAATAAACCAAGACTT 590

QY 121 AlaLeuSerLeuGlyValAlaLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140  
 DB 591 GCTTTAAGCCCTGGGAGTGGGCTTATCATTTAGCTTACAGCTGTGAGACAGTAAAAAAG 650  
 QY 141 ProPheIleProIleHisHisMetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
 DB 651 CCATTCAATTCCTCATTCATATGAGAGCTCATGACTATTAATGATGACCAATATAA 710  
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
 DB 711 GTAGAAATTCCTTTTATGTTCTTTTGATTTCTGAGGTCACTGCTGTGGCATTAAT 770  
 QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
 DB 771 CAAGAGTTTACAGATTTCTGCTCTTGGAAAGTCTTTGACATGACACAGGTGACATG 830  
 QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220  
 DB 831 CTGACAAAGTGGCAAGAAAGACTTTCTTAATAAACAATCCAGAGTGTCCACCAATGAGT 890  
 QY 221 GlyGlyLysValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
 DB 891 GGTGGAAAGCCATPAGAACATTTGGCCAAACAGGAATAGATTTCAATTTGACATCAAA 950  
 QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
 DB 951 CTCCCTTCATCATGCTAAATAATTTGATTTCTTTTACGTGACCTCAACAGCTTACT 1010  
 QY 261 AspLysIleIleMetLysLysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSer 280  
 DB 1011 GATAAATATATATGAAATAAGAAAAAGCAAGGATATGGAAGGGCCAAATCCTGCT 1070  
 QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisSerMetAlaCysHisLeuValLysArg 300  
 DB 1071 TCAGACAGACGACATTTGCTCCACAGTACAGCAACAATGCGATGCTGTGTGAAAAAGA 1130  
 QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320  
 DB 1131 ACAATCGGCTATTTCTGTTTGTGACAGAGAGACTTTTACTCCAAATATATGACATG 1190  
 QY 321 LeuValAlaSerGlyValAlaAlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeu 340  
 DB 1191 CTGGTTGATCTGGTGGTGTGCAAGTAACTTCTATATCCGACAGCTCTGGAATTTTA 1250  
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaGluLeuCysThrAspAsnGly 360  
 DB 1251 ACAACGCAACACAGTGCATTTGTGTCTCTCCACAGACTATGACATGATATATGTC 1310  
 QY 361 IleMetIleAlaTTPAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
 DB 1311 ATTATGATTTGATGAAATGATTAAGACTACCTGCTGTGGCATTTTATCATGAC 1370  
 QY 381 IleGluGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerLysGluVal 400  
 DB 1371 ATGAAGGATCCGCTATGAAACCAAAATGTCTCTTGAGATGACATATCAAAAGAGTT 1430  
 QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414  
 DB 1431 GGAGAGCTTCATTAAGTACCAATTAATAATGAGATA 1472  
 RESULT 3  
 AR428808 AR428808 1387 bp DNA linear PAT 18-DEC-2003  
 LOCUS Sequence 21 from patent US 6642041.  
 DEFINITION AR428808  
 ACCESSION AR428808  
 VERSION AR428808.1 GI:40188594  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 1387)  
 AUTHORS Chen, J., Feder, J.N., Nelson, T.C., Krystek, S.R. and Duclos, F.



TITLE Polynucleotides encoding a novel metalloprotease, MP-1  
JOURNAL Patent: US 6642041-A 21 04-NOV-2003;  
FEATURES Location/Qualifiers  
source 1..1387  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN

Alignment Scores:

| Pred. No.:             | 3,54e-168 | Length:       | 1387 |
|------------------------|-----------|---------------|------|
| Score:                 | 2090.50   | Matches:      | 412  |
| Percent Similarity:    | 93.85%    | Conservative: | 0    |
| Best Local Similarity: | 93.85%    | Mismatches:   | 2    |
| Query Match:           | 98.38%    | Indels:       | 25   |
| DB:                    | 6         | Gaps:         | 1    |

US-10-649-273-2 (1-414) x AK428808 (1-1387)

QY 1 MetLeuIleLeuThrIysThrAlaGlyValPhePheIysProSerIysArgIysValItyr 20  
DB 24 ATGCTATATCTTGACTAGACTGACGAGGTTTATTAACATCAAAAGGAAGTTAT 83  
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIlysiIleValLeu 40  
DB 84 GAATTTTAAAGAGTTTAAATTTTCACTCCGAAACACTATTTCTTCAATAATAGTATTG 143  
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAspGluThrGlyAsnVal 60  
DB 144 GGAATTGAACCTGATGTGTGATGATGACGAGCTGTGTGTGATGAACTGGAAAGTGTG 203  
QY 61 LeuGlyGluAlaIleHisSerGlnThrGlyValHisIleuThrGlyIlyIleValPro 80  
DB 204 TTGGGAGAACATATCACTTCCCAAGCTGAAGTTCAATTAATAACAGGTGGATTGTCT 263  
QY 81 ProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnIleValLeuSer 100  
DB 264 CCAGGAGCTCAACAGCTTCAAGAGAAATATTCAAGAAATAGTACAAAGCTTTCT 323  
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleIysProGlyLeu 120  
DB 324 GCACAGTGAAGTCTCTCCAAAGTCACTTCAAGCAATTCACATCAATTAACAGGACTT 383  
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyIleLeuIysLys 140  
DB 384 GCTTTAAGCTCGGAGAGGCGCTTATCTTATAGCTTACAGCTGAGACAGTTTAAAG 443  
QY 141 ProPheIleProIleHisIleSmeGluAlaHisIleAlaLeuThrIleArgLeuThrAspLys 160  
DB 444 CCATTCATTCCTCATTCATCATGAGAGCTCATGCACTTATAGTTGACCAATATAA 503  
QY 161 ValGluPheProPheLeuValIleLeuIleSerGlyIlyHisIleCysLeuLeuAlaLeuVal 180  
DB 504 GAGAATTTCTCTTTTATGTTCTTTGATTTCTGAGAGTCACTGTCGTTGGCATTAGTT 563  
QY 181 GlnGlyValSerAspPheLeuLeuGlyIlySerLeuAspIleAlaProGlyIysAspMet 200  
DB 564 CAAAGAGTTTCAATTTTCTGCTTCTTGAAGCTTTTGAACATAGACCAAGGAGCATG 623  
QY 201 LeuAspIlyValAlaArgArgLeuSerLeuIleIysHisProGlyCysSerThrMetSer 220  
DB 624 CTTGACAGGTGCGAAGAGACTTTCTTTAATAAACAATCCAGAGTCTCCACCATAGAGT 683  
QY 221 GlyIlyIlyValAlaIleGluHisIleLeuAlaIlyGlnGlyAsnArgPheHisIlyAspIleIys 240  
DB 684 GGGGAGGAAAGCCATAGACATTTGGCCAAACAGAAATAGATTCTTTTGAATCATCAAA 743  
QY 241 ProProLeuHisHisIleAlaIlyAsnCysAspPheSerPheThrGlyLeuGlnHisIleValThr 260  
DB 744 CCGCCCTTGATCATGCTAAATAATGTGATTTTCTTTTACGTGACTTCAACAGTTACT 803  
QY 261 AspIlyIleIleMetIlyIlyGluIlyGluIlyIle----- 273  
DB 804 GATTAATAATATATGAAAGGAAAGGAAAGGAGTATATTCTTAATTAGTAAAGTTGAA 863

QY 274 -----GluLys 275  
DB 864 CAGTAAATATTCCTGATGTGCTTAAATAATAGCTCATTTCTGACGAGTAAGAGAAG 923  
QY 276 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 295  
DB 924 GGGCAATTCCTGCTTCAAGCAGACATCTCTCCACATGACGACACAAATGCGCATGT 983  
QY 296 HisIleuValIysArgThrHisArgAlaIleLeuPheCysIlyGlnArgAspLeuLeuPro 315  
DB 984 CATTTGTGAAGAAACACATCGGCTATTTCTGTTGTGAAGCAGAGACTGTTTACT 1043  
QY 316 GlnAsnAsnAlaValIleuValAlaSerGlyIlyValAlaIleSerAsnPheTyrlleArgArg 335  
DB 1044 CAAATATATGACAGTACGTGTTGACATCTGCTGTGTGCGCAATTAATCTTATATCCGACA 1103  
QY 336 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 355  
DB 1104 GCTCTGAAATTTTAAACAAACGACACAGTCACTTTGTTGTGTCTCTCCAGACTA 1163  
QY 356 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeu 375  
DB 1164 TGCACTATATATGCAATTAATGATGATGAAATGATATGAAAGACTACGTTGCTTG 1223  
QY 376 GlyIleLeuHisAspIleGluGlyIleArgTyrgluProLysCysProLeuGlyValAsp 395  
DB 1224 GGCATTTTACATGACATAGAGGCAATCCGCTATAGAACCAAAATGTCTCTTGAGTACAC 1283  
QY 396 IleSerIysGluValGlyGluAlaSerIleIysValProGlnLeuIysMetGluIle 414  
DB 1284 ATATCAAAAGAGTTGGAGAAAGCTTCCATTAAGTACACATTAATAATGAGATA 1340  
  
RESULT 4  
HSA295148 1387 bp mRNA linear PRI 30-OCT-2000  
LOCUS Homo sapiens mRNA for putative staloglycoprotease type 2.  
DEFINITION A0295148  
ACCESSION A0295148.1 GI:11071726  
VERSION A0295148.1 GI:11071726  
KEYWORDS metalloproteinase; staloglycoprotease.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
  
REFERENCE  
1 Chen,J.M., Fortunato,M. and Barrett,A.J.  
Cloning and sequencing of a second human putative  
staloglycoprotease homologue  
Unpublished  
2 (bases 1 to 1387)  
Chen,J.M.  
Direct Submission  
Submitted (27-OCT-2000) Chen J.M., MRC Molecular Enzymology  
Laboratory, The Babraham Institute, Babraham, Cambridge, CB2 4AT,  
UNITED KINGDOM  
  
FEATURES  
source Location/Qualifiers  
1..1387  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/tissue\_type="uterus"  
24..1343  
/function="metalloproteinase of family M22"  
/codon\_start=1  
/product="putative staloglycoprotease type 2"  
/protein\_id="CAC14666.1"  
/db\_xref="GI:11071727"  
/db\_xref="GOA:Q9H4B0"  
/db\_xref="UniProt/TREMBL:Q9H4B0"  
/translation="MLILTKAGVFFKPSKKVVEFLRSFNFHEETLFLAKIVGIEIT  
SCDPTAAVVDENGVGEAIIHSQTEVLAIKTKGVLPAPAOQHSENIORIVQEAISAS  
GVSPSDLSAIAITTKIGLAISGVLSFSLQVQAKKRPPIRHHMEHALITLTK  
VEPFLVLLISGHCILALVQGVSDPFLILGSLDIAPQMDLAKARRLSLIKHECST



```

QY 41 GYIILEGIuThrSerCysAspAspThrAlaAlaValAlaAspGluThrGlyValSerVal 60
DB 121 GGAAATTTGAACTAGTGTGTGTATGATACAGCAGCTGTGTGTGTATGAAATCGGAAATGTC 180
QY 61 LeuGlyValAlaAlaIleHisSerGlnThrGluValHisIleuLeuThrGlyValIleValPro 80
DB 181 TTGGGGAAAGCATACTTCCCAACTGAGAGTTCACTTTAAACAGGTGGGATTTGTTCT 240
QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
DB 241 CCAGGAGCTCAACAGCTTCACAGAAATATTCACAGAAATAGTACAGAGAGCTTTCT 300
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
DB 301 GCCAGTGGAGTCTCTCCAGTGTACCTCTCAGCAATTCACATCAATTAACAGAGACTT 360
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140
DB 361 GCTTTAAGCCTGGAGAGTGGCTTATCATATTAGCTTACAGCTGTATGACAGTTAAAG 420
QY 141 ProPheIleProIleHisIleHisMetGluAlaHisIleAlaLeuThrIleArgLeuThrAsnLys 160
DB 421 CCATTATTCCTCATTCATCATATGAGAGCTCATGCACTTACTATTTAGGTTGACCAATTA 480
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyValHisCysLeuLeuAlaLeuVal 180
DB 481 GTAGAAATTCCTTTTATGTTTCTTTGATTTCTTGAGAGTCACTGTCTGTTGGCATTTAGT 540
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
DB 541 CAAGGAGTTTCAGATTTTCTGCTTCTTGAAAGTCTTTGACATATGACACAGAGTGACATG 600
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisIleProGluCysSerThrMetSer 220
DB 601 CTTCGACAGGTGGCAGAGACATTTCTTTAATAAACATCCAGAGTCTCCACCATAGT 660
QY 221 GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
DB 661 GGTGGGAAAGCCATAGAACATTTGGCCAAACAGAAATGATTTTCACTTTTGGACATCAA 720
QY 241 ProProLeuHisIleAlaLysAsnGlyAspPheSerPheThrGlyLeuGlnHisValThr 260
DB 721 CCTCCCTTCATCATCTATAAAATTTGATTTTCTTTTACCTGACTTCAACAGTTACT 780
QY 261 AspLysIleIleMetLysGlyLysGluGlnGlyIleGlnLysGlyGlnIleLeuSer 280
DB 781 GATTAATAATATGATGAAACAGGAAACAGAGAAAGTATTAAGAAAGGGGCAAAATCTGTCT 840
QY 281 SerAlaAlaAspIleAlaIleAlaThrValGlnHisIleThrMetAlaCysHisLeuValLysArg 300
DB 841 TCAGCAGCAGACATTTCTGTCCACAGTACAGCACAATGGCATGTCTCTTGTGAAAGA 900
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320
DB 901 AACACATCGGAGCTATTCGTTTCTTAAGCAGAGAGACTTGTACTCAAAATATATGCAGTA 960
QY 321 LeuValAlaSerGlyValAlaAlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeu 340
DB 961 CTGGTTCATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1020
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360
DB 1021 ACNAAAGCACAACAGTCACTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1080
QY 361 IleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
DB 1081 ATTATGATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
QY 381 IleGluGlyIleLeuArgGluProLysCysProLeuGlyValAspIleSerLysGluVal 400
DB 1141 ATAGAAGGATCGCTATGAAACCAAAATGTCTCTTGAGATAGACATATCAAAAGAGTT 1200
QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414

```

```

DB 1201 GGAAAGCTTCATAAAATGATCCAAATTAATAATGAGATA 1242
RESULT 6
LOCUS AX664695 1820 bp DNA linear PAT 22-MAR-2003
DEFINITION Sequence 4 from Patent WO02074960.
ACCESSION AX664695
VERSION AX664695.1 GI:29164455
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 Leiby,K.R., Kapeller-Liberman,R. and Glucksmann,M.
AUTHORS 38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions
TITLE of human proteins and uses thereof
JOURNAL Patent: WO 02074960-A 4 26-SEP-2002;
Milleium Pharmaceuticals, Inc. (US)
FEATURES
source
1..1820
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
146..1390
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD80044.1"
/db_xref="GI:29164456"
/translacion="MLITKTAIVFEPKSRKRYEFLRSFNHPTLFLKIVIGIET
SCDDTAAAVDENGVNVEBAIHSQTEVHLKTGGIVPPAAQOIHRENIQRIVOEALNS
VSPSDLSAATTKIPGLASLGSFSLQVGLKPEPLPIHREMAHLITLTK
VERPPLVILISGHCILAIQVQSDPLFLKSLIDAPGMDLKVAREPLSLIKPEGST
MSGKAIENLAKGNRFPDIKPPLEHAKCDSFTGLQYTTDKNNRQBESEIKG
QILSSADINATVQHTWACHLVKRTKRAIILFCQQRDLPPNNAVLVASGVASNPYLR
RALEIITNATQCTLCPPRLCTDNGIMIMNGIERLRAGILHDIIGIRYBKCPL
GVDSIKVGBASIKVPOLKMEI"
ORIGIN
Alignment Scores:
Pred. No.: 8.14e-168 Length: 1820
Score: 2088.00 Matches: 407
Percent Similarity: 99.03% Conservative: 3
Best Local Similarity: 98.31% Mismatches: 4
Query Match: 98.26% Indels: 0
Gaps: 0
DB: 6
US-10-649-273-2 (1-414) x AX664695 (1-1820)
QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20
DB 146 ATGTAATCTTGACTTAAGACTGACGAGGATTTTAAACCATCAAAAGGAAAGTTTAT 205
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrPhePheLeuHisLysIleValLeu 40
DB 206 GAATTTTAAAGAGTTTAAATTTTCACTCTCGAACAATATTTCTTCAATAAAATGATTTG 265
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAlaValAspGluThrGlyValSerVal 60
DB 266 GAATTAACACTAGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 325
QY 61 LeuGlyValAlaIleHisSerGlnThrGluValHisIleuLysThrGlyValIleValPro 80
DB 326 TTGGGAAACCAATACATTTCCCAACTGAAGTTCAATTAATAACAGGTGGGATGTTCTT 385
QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
DB 386 CAGACACTCAACAGCTTCACAGAGAAATATTTACAGAAATAGTACAGAGAGCTTTTCT 445
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
DB 446 GCGAGTGGAGTCTCTCCAGTGTACCTCTCAGCAATTCACATCAATAAAACAGAGACTT 505

```

QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140  
 DB 506 GCTTTAAGCTGGAGAGTGGGCTTATCTTAGCTTACAGCTGTAGACACATTAAG 565  
 QY 141 ProPheIleProIleHisIleMetGlnLysAlaIleLeuThrIleArgLeuThrAsnLys 160  
 DB 566 CCATTCACTCCCATTCATCATATGAGAGCTCATCTACTATTAGGTTGACCAATTA 625  
 QY 161 ValGlnPheProPheLeuValIleLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
 DB 626 GTGAATTCCTCTTTTATGTTCTTTGATTTCTGAGGTCACTGCTGTTGCACTTACT 685  
 QY 181 GlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
 DB 686 CAAGAGCTTCAGATTTTCTGCTCTTGGAAGCTTTTGACATAGACACAGAGTACATG 745  
 QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyCysSerThrMetSer 220  
 DB 746 CTGACAAAGGTGGCAAGACCTTTCTTAATAAATCAATCCAGAGTCTCCACCATGAGT 805  
 QY 221 GlyGlyValAlaIleGlyHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
 DB 806 GGTGGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATTTCATTTGACATCAAA 865  
 QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
 DB 866 CTTCCCTTCATCTAGCTTAAATAATTTGATTTTCTTTACTGACATTCACACATTAAT 925  
 QY 261 AspLysIleIleMetLysLysGlyLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 280  
 DB 926 GATTAATAATATGAAACAGAAACAGAAAGAGATATGAGAGGAGCAATCTGCT 985  
 QY 281 SerAlaAlaAspIleAlaIleAlaIleThrValGlnHisThrMetAlaCysHisLeuValLysArg 300  
 DB 986 TCAGACACACATTCGTCGACACAGATACAGACACAAATGCAATGCTCATCTTGTGAAAAGA 1045  
 QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGlnAsnAsnAlaVal 320  
 DB 1046 ACACATGGGCTATCTGTTTGTGTAAGCAGAGACTTGTATCTCAAAATATATGCATTA 1105  
 QY 321 LeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeu 340  
 DB 1106 CTGCTTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1165  
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysAspProProArgLeuCysThrAspAsnGly 360  
 DB 1166 ACAAGCAACACAGTGCATTTGTGTCTCTCCCAAGCTATGCATGATATATGCT 1225  
 QY 361 IleMetIleAlaTyrAsnGlyIleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
 DB 1226 ATATATGATTCATGATATGATATGATATGATATGATATGATATGATATGATATG 1285  
 QY 381 IleGlnGlyIleArgTyrGlnProLysCysProLeuGlyValAspIleSerLysGlyVal 400  
 DB 1286 ATAAAGGCAATCCCTATGAACCAAAATGTCTCTTGGAAGTATGACATATCAAAAGAGTT 1345  
 QY 401 GlyGlyLysSerIleLysValProGlnLeuLysMetGlnIle 414  
 DB 1346 GGAAGACCTTCATTAAGTACCAATTAATAAATGAGATTA 1387  
 RESULT 7  
 AX713716 2208 bp DNA linear PAT 15-APR-2003  
 LOCUS Sequence 400 from Patent EP1293569.  
 DEFINITION AX713716  
 ACCESSION AX713716.1 GI:29888642  
 VERSION  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

AUTHORS Isegai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,  
 Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,  
 Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and  
 Masuko, Y.  
 TITLE Full-length cDNA  
 JOURNAL Patent: EP 1293569-A 400 19-MAR-2003;  
 Helix Research Institute (JP); Research Association for  
 Biotechnology (JP)  
 FEATURES  
 source location/Qualifiers  
 1..2208  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,83e-155 Length: 2208  
 Score: 1944.00 Matches: 386  
 Percent Similarity: 93.24% Conservative: 0  
 Best Local Similarity: 93.24% Mismatches: 4  
 Query Match: 91.48% Indels: 24  
 DB: 6 Gaps: 1  
 US-10-649-273-2 (1-414) x AX713716 (1-2208)  
 QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20  
 DB 344 ATGCTAATCTTGACATCAAGACATGCAAGAGTTTAAAAACATCAAAAAAGAAAGTTTAT 403  
 QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40  
 DB 404 GAATTTTAAAGAGTTTAAATTTTATTCATCTGGAACATATTTCTCATTAATAATGATATG 463  
 QY 41 GlyIleGlnThrSerCysAspAspThrAlaAlaValAlaValAspGlnThrGlyLysVal 60  
 DB 464 GGAATTTAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 523  
 QY 61 LeuGlyGlyAlaIleHisSerGlnThrGlnValHisLeuLysThrGlyLysIleValPro 80  
 DB 524 TTGGAGAAAGCAATACATCTCCAACTGAAAGTTCAATTTAAACAGAGTGGATTTTCT 583  
 QY 81 ProAlaAlaGlnGlnLeuHisArgGlyLysAsnIleGlnArgIleValGlnGlnAlaLeuSer 100  
 DB 584 CCAGAGCTCAACACCTTCAAGAGAAATATTAACGAATGTCACAGAACTCTTCT 643  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
 DB 644 GCCAGTGAAGTCTCTCAAGTGAATCTCTCAGCAATTCGAACTTACCAATTAACCAAGACTT 703  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140  
 DB 704 GCTTTAAGCTGGAGTGGGCTTATCATTTTACCTTACAGCTGTGAGACAGTTAAAG 763  
 QY 141 ProPheIleProIleHisIleMetGlnLysAlaIleLeuThrIleArgLeuThrAsnLys 160  
 DB 764 CCATTCACTCCCATTCATCATATGAGAGCTCATGCACTTACATTAAGTTCAGCAATAAA 823  
 QY 161 ValGlnPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
 DB 824 GTGAATTCCTCTTTTATGTTCTTTGATTTCTGAGAGTCACTGCTGTGCAATTAAGT 883  
 QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
 DB 884 CAAGAGTTCAGATTTTCTGCTCTTGGAAGTCTTTGAGATATGACACAGATGACATG 943  
 QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyCysSerThrMetSer 220  
 DB 944 CTGACAAAGTGGCAAGAGACTTCTTAATAAACAATCCAGAGTCTCCACATGAGT 1003  
 QY 221 GlyGlyLysAlaIleGlyHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
 DB 1004 GGTGGGAAAGCCATAGAACATTTTGGCCAAACAGAAATATGATTTCAATTTGACATCAAA 1063

|            |  |             |   |                 |
|------------|--|-------------|---|-----------------|
| Qy         |  | 241         | ProProLeuHiSHSAlaIyaAsnCyaspPheSerPheThGlyLeuGlnHisValThr       | 260             |
| Db         |  | 1064        | CCTCCCTTGACATCAGCTCAAAAATTGTATTTTCTTTACTGCAGCTTCACACCGTACT      | 1122            |
| Qy         |  | 261         | AspLySIlelleMeLysLysGIuLysGIuLuglylilegluylsgilyelnlleuSer      | 280             |
| Db         |  | 1124        | GATTAATAATATATCATGAAGAAGAAAAAGAGAAAGATATTGAGAAAGGGCAATCTGTCT    | 1183            |
| Qy         |  | 281         | SerIlaIlaaspiIIeAlaIaIaThryValGlnHISthrMetAlaCyHisIleuVallysArg | 300             |
| Db         |  | 1184        | TCAACAGACGACATTTGCTGCCACAGATACAGACAACAATGGCAATGCATCTTGTGTAAAGA  | 1243            |
| Qy         |  | 301         | ThriIsARgAlaIleLeuPheCyelsySGInArGaSPLeuLeuProGIAspaSnalAval    | 320             |
| Db         |  | 1244        | ACAAGTCGGGCTATTCTCTTTTGTAGACAGAGACTGTTTACCCTCAAAATATACAGTA      | 1307            |
| Qy         |  | 321         | LeuValAlaSeRgiYGIyValAlaSeXanPheTyrlleaRgaGalaleuGluleu         | 340             |
| Db         |  | 1304        | CTGGTTGCATCTGGGGTGTGCGCAAGTAATCTTGTATCCGACAGCTCTGGAAATTTTA      | 1363            |
| Qy         |  | 341         | ThrsnaIatTrngInCySThrLeuLeuCysProPProIaRgleuCySThrAspaNgly      | 360             |
| Db         |  | 1364        | ACAACCGAACACACAGTGACTTGTGTGCTCTCCACAGCTATGCACTGATATATGCG        | 1423            |
| Qy         |  | 361         | IleMetIleAlaTPspaNglylilegluarGuenaRgalaglyLeuGlylileuHiSasp    | 380             |
| Db         |  | 1424        | ATTATGATTGCA-----   | 1433            |
| Qy         |  | 381         | IlegluGlyIleaRgyRGluProLysCysProLeuGlyValAspIleSerLysGIuVal     | 400             |
| Db         |  | 1436        | -----TGAGTCTCTCTTGAGTAGTAGCATATCCAAGAAAGTTT                     | 1471            |
| Qy         |  | 401         | GIyGLuaIaseRIleLysValProGIuLeuLysMetGIulle                      | 414             |
| Db         |  | 1472        | GGAGAAGCTTCATATAAAAGTACCACAATATAAAATGAGAGATA                    | 1513            |
| <hr/>      |  |             |   |                 |
| RESULT 8   |  |             |   |                 |
| AKO55441   |  |             |   |                 |
| LOCUS      | AKO55441   | 2208 bp     | mRNA  | linear          |
| DEFINITION | Homo sapiens cDNA FLJ30879 fis, clone FEBRA2004592, highly similar |             |   | PRI 30-JAN-2004 |
| ACCESSION  | AKO55441   |             |   |                 |
| VERSION    | AKO55441.1   | GI:16550166 |   |                 |
| KEYWORDS   | oligo capping, fis (full insert sequence).                         |             |   |                 |
| SOURCE     | Homo sapiens (human)   |             |   |                 |
| ORGANISM   | Homo sapiens   |             |   |                 |

REFERENCE  
AUTHORS

1  
Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,  
Wakamatsu, A., Hayashi, K., Sato, H., Negai, K., Kimura, K., Makita, H.,  
Sekine, M., Odayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,  
Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,  
Nagahori, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,  
Shiratori, A., Sudo, H., Hosoi, T., Kaku, Y., Kodaira, T., Kondo, H.,  
Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,  
Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,  
Tanikawa, M., Yamazaki, M., Niinomiya, K., Ishibashi, T., Yamashita, H.,  
Muraoka, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,  
Hiraoaka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,  
Yosida, M., Hottaru, T., Kusano, J., Kanehori, K., Takahashi-Fuji, A.,  
Hata, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R.,  
Takeuchi, K., Arita, M., Imose, N., Mueshino, K., Yuki, H., Oshima, A.,  
Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,  
Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Saich, N., Takami, S.,  
Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,  
Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,  
Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,  
Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,  
Tashiro, H., Tanigami, A., Fujitani, T., Ono, T., Yamada, K., Fujii, Y.,  
Ozaki, K., Hiroo, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,  
Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,  
Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsunuma, K.,

TITLE  
JOURNAL  
PUBLISHED  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
PUBLISHED  
REFERENCE  
AUTHORS

COMMENT

FEATURES  
source

CDS

ORIGIN

Alignment Scores:  
Pred. No.:  
Score:  
Percent Similarity:  
Best Local Similarity:  
Query Match:  
DB:

US-10-649-273-2 (1-414) x AK055441 (1-2208)

1 MetleuileuthrlystrnralagiValPhehPhelysProserlysaRglysValIyr 20  
Db 344 ATGTAATCTTGAACCTGACGAGGAGTTTATTTTAAACATCAAAAAGAAAGTTTAT 403

21 GlupheleuArygserPheasnPhenHsPrglyThrlleuPhelenuHslylsileValleu 40

1.83e-155 Length: 2208  
1944.00 Matches: 386  
93.24% Conservative: 0  
Percent Similarity: 93.24% Mismatches: 4  
Best Local Similarity: 91.48% Indels: 24  
Query Match: 91.48% Gaps: 1  
DB: 9

location/Qualifiers  
1..2208  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="FEBRA2004592"  
/tissue\_type="brain"  
/clone\_lib="FEBRA2"  
/dev\_stage="fetus"  
/note="cloning vector: pME18SEFL3"  
344..1438  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAB70923.1"  
/db\_xref="GI:16550167"  
/translation="MLLTITAGVFPSPKRYTERPSRNFHNGTLFLKIVLGIT  
SCDUTAAVAVDETGVGEAHSQTEVHLTGTGIVPPAAQQLHEENTQIVQELASLS  
GVSDSDLSAATATTIKPGLASLSTGSLFSQTLVGQALKPFIPIHMEAHALTITLTK  
VSEPFVILITSGCHLALVQGVSDFFLLKSLDIAGMDLKVARYLPIIKHPCST  
MSEGGKALEHAKCOGNPHFDIKPILHAKKCDPFGTGLQVTDKIKKEKESEGIKGG  
QIISASDIATAOHTMACGLVYKTRHAILFCQRDLDPNNMVLVASGVASNFCIR  
RALEILTNATQCTILCPPLRLTDNGMIMA"

3 (bases 1 to 2208)  
Iisogai,T., Otsuki,T. and Sugiyama,T.  
Direct Submission  
Submitted (24-OCT-2001) Takao Iisogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail:genomics@hi.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
NEBO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: RAB and  
HRI.

|    |      |  |      |
|----|------|--|------|
| Db | 404  | GAATTTTAAAGAGTTTAAATTTCATCTCTGGAAACACTATTCTTCATAAATAGTATGG     | 463  |
| Oy | 41   | GIYIIEGIUTRISerCyAspAspThrLalalaValValAspGIUTRGIYAspVal        | 60   |
| Db | 464  | GGAAATTGAACCTAGTGTGTGATGATACACACACTCTGTGTGTGATGAACCTGGAAATGTG  | 523  |
| Oy | 61   | LeuGIYIuaIaIeHISerGIUTRGIYValHISleuLysTRNGIYIIEValPro          | 80   |
| Db | 524  | TTGGAGAAAGCAATTCATTTCCCAACTGAAGTTCATTTAAAACAGGGGGATGTCTCT      | 583  |
| Oy | 81   | ProAlaIaGIuNGInLeuHISArgGIuAsnIleGIuArgIIEValGIuAlaLeuSer      | 100  |
| Db | 584  | CCAGAGCTCAACACACTTCACAGAGAAATATTCAAGAAATAGTACAAAGAGCTCTTCT     | 643  |
| Oy | 101  | AlaSerGIYValSerProSerAspLeuSerAlaIlealalThrTrIIElyProGIYLeu    | 120  |
| Db | 644  | GCCAGTGAAGTCTCTCCAAAGTGAAGCTTCACACAAATGGCACTTAAACCAAGACTT      | 703  |
| Oy | 121  | AlaLeuSerLeuGIYValGIYLeuSerPheSerLeuGIuValGIuLeuLysLys         | 140  |
| Db | 704  | GCTTTAACTCTGGAGTGGGCTTATCATTTAGCTTACAGCTGTGTGTGTGTGTGTGTGT     | 763  |
| Oy | 141  | ProPheIIEProLIEHISISemecGIuAlaHISAlaLeuTrIIEArgLeuThrAspLys    | 160  |
| Db | 764  | CCATTCAATTCATTCATTCATATGAGAGCTCATGACTTACTATTAAGTTGACCAATAA     | 823  |
| Oy | 161  | ValGIuPheProPheLeuValLeuLeuIIEserGIYGIYHISCyuValLeuAlaVal      | 180  |
| Db | 824  | GTAGAAATTTCTTTTGTATGTTCTTTGTATTTCTGAGAGCTCACTGTCTGTGTGTGT      | 883  |
| Oy | 181  | GIuNGIYValSerAspPheLeuLeuLeuGIYLysSerLeuAspIIEalalProGIYAspMet | 200  |
| Db | 884  | CAAGGAGTTCCAGATTTCTGTCTTTGGAAAGCTTTGTACATATGACACAGAGTGAATG     | 943  |
| Oy | 201  | LeuAspLysValAlaArgArgLeuSerLeuIIElySHISProGIYCySerThrMetSer    | 220  |
| Db | 944  | CTTGACAAAGTGGCAAGAAAGCTCTTTATTAACAATCCAGAGTCTCCACATAGAT        | 1003 |
| Oy | 221  | GIYGIYLYsaIaIIEGIuHISleuAlaLysGIuNGIYAsnArgPheHISAspIIElyS     | 240  |
| Db | 1004 | GGTGGAAAGCCATGAGACATTTGGCCAAACAGGAATTAAGATTTCATTTGACATCAAA     | 1063 |
| Oy | 241  | ProProLeuHISAlaLysAsnCyAspPheSerRhetNGIYLeuGIuHISValThr        | 260  |
| Db | 1064 | CCTCCCTTCACATGACTTAAATGTGATTTCTTTTACTGACATTCACACAGTACT         | 1123 |
| Oy | 261  | AspLysIIEIIEmetLysLysGIuLysGIuNGIYIIEGIuLysGIYAlIIEleuSer      | 280  |
| Db | 1124 | GATTAATAATATATGATAAAGAAAGAAAGAAAGGATATTGAAAGGGGCAATCTGTCT      | 1183 |
| Oy | 281  | SerAlaAlaAspIIEalalalThrValGIuHISThrMetAlaCyuHISleuValLysArg   | 300  |
| Db | 1184 | TCAGCAGACGACATTTGCTGCCACAGTACACACAAATGGAGTGCATCTGTGTAAAGA      | 1243 |
| Oy | 301  | ThrHISArgAlaIIEleuPheCysLysGIuArgAspLeuLeuProGIuAsnAsnAlaVal   | 320  |
| Db | 1244 | ACACATCGGCTAATCTGTTTGTGAAGCAGAAAGCTTGTAACCTCAAAATATATGCAC      | 1303 |
| Oy | 321  | LeuValAlaSerGIYGIYValAlaSerAsnPheTrIIEArgArgAlaLeuGIuIIEleu    | 340  |
| Db | 1304 | CTGTGTGCATCTGTGTGTGTGTGCAGATAACTTCTGTATTCGGAGAGCTCTGGAAATTT    | 1364 |
| Oy | 341  | ThrAsnAlaThrGIYCyThrLeuLeuCyProProArgLeuCySerThrAspAsnGIY      | 360  |
| Db | 1364 | ACAAACGCAACACAGTGCATTTGTGTGTCTCTCCAGACTATGACATATATGGC          | 1422 |
| Oy | 361  | IIEmetIIEalalTPAsnGIYIIEGIuArgLeuArgAlaGIYLeuGIYIIEleuHISAsp   | 380  |
| Db | 1424 | ATTATGATTCGCA-----   | 1425 |
| Oy | 381  | IIEGIuGIYIIEArgTrGIuProLysCyProLeuNGIYValAspIIEserIIEGIYVal    | 400  |
| Db | 1436 | -----TGATTCCTCTTTGAGATGAGCAATATCAAAAGACT                       | 1471 |

[illegible]

source

1. 1844  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6 J"  
/db\_xref="taxon:10090"  
/clone="MGC:67870 IMAGE:5012054"  
/isue\_type="Mammary tumor metastatized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."  
/clone\_1ib="NCI CGAP\_Lu29"  
/lab\_host="DH10B"  
/note="Vector: PCMV-SPORT6"  
222. 1466  
/product="Unknown (protein for MGC:67870)"  
/protein\_id="AA58172.1"  
/db\_xref="GI:34849664"  
/translation="MLMRLRTAGAIIPKPKSKYVGLRFRSVHPRILTSCHKLVIGLET  
SCDDTAAVVDENYVGEALHSQYVHLRTGIVPVAQOHLRENTORIVEETLSAS  
RTPSDLSAATTIKPGLALSLGVLSFSLQVNRFPKPIPIHMEHALITRLTK  
VEPFLVLIISGHCLLALVQVSDFLKSLDIAGMDLVARLSLIRKDECT  
MEGKALIEHLAKDGNRPHFTINPMONAKKCDSPFCGLITDKLTHKEKEGEMG  
QLSSAADIYAAVQHTACHLAKTRHAILFCOKILSPANAVLVSGVASLTYR  
KALEIVANAIOCTLLCPPEPLCTDNGIMAMNGIERLAGLGLHVEDIRYEPKPL  
GDISREVAEAALKVPRKXAL"  
333. 1397  
/note="ORF7: Region: Metal-dependent proteases with possible chaperone activity [posttranslational modification, protein turnover, chaperones]"  
/db\_xref="CCD:COG0533"

misc\_feature

Alignment Scores:

Pred. No.: 2, 71e-146 Length: 1844  
Score: 1835.00 Matches: 352  
Percent Similarity: 91.06% Conservative: 25  
Best Local Similarity: 85.02% Mismatch: 37  
Query Match: 86.35% Indels: 0  
DB: 10 Gaps: 0

US-10-649-273-2 (1-414) x BC058172 (1-1844)

QY 1 MetLeuileuThrIysThrAlaGlyValPhePhePheProSerIysArgLysValIyr 20  
DB 222 ATCTAATGTTAAGAGACAGACAGAGCTATCCCAAGCCCCCAAGAGTAAGTTAT 281  
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysValIeu 40  
DB 282 GGAATTTTAAAGAGATTAGTTCATCCAGAACTCTCTTGTCTAATAAGTCTCTG 341  
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60  
DB 342 GGAATTTTAAAGAGATTAGTTCATCCAGAACTCTCTTGTCTAATAAGTCTCTG 401  
QY 61 LeuGlyValAlaIleHisSerGlnThrGluValHisLeuLysThrGlyValIeuPro 80  
DB 402 CTGGGGGAGACACTGCTCCCAACTCAGGTCATCTGAAACAGGTTGGATTTGTC 461  
QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
DB 462 CCAGTAGCTCAACACTTCACAGAAATATTCAGAAATAGTAGAAGAACTTTCT 521  
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
DB 522 GCGCAGTAGAATCACCCCAACGATCTCTCGCAATTCACATCAACACCGGAGCTG 581  
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnIleValGlyIleLeuLys 140  
DB 582 GCGCTAAGCTTAGAGAGCTTATCTTAGCTTAGACGTAGTAAGTCCGTTTAAAG 641  
QY 141 ProPheIleProIleHisIsmetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
DB 642 CCGTTATCCGATTCATCATGAGAGCTCACGCACTGACTATTAGGCTCACCAATRAA 701

QY 161 ValGluPheProPheLeuValIleuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
DB 702 GTAGAAATTTCTTTTCTTTAGTTCTTTTATTTCTGCGGCTCATCTGTTGGCATTTAGT 761  
QY 181 GlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
DB 762 CAAGGTCTTTCCGATTTCTGCTCTCTGGGAAAGTCTTTGGACATAGCGCGGAGCATG 821  
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleHisProGlyCysSerThrMetSer 220  
DB 822 CTGACAAAGGTGGCAAGAGCTTTCTTTATTCAAACATTCAGAAATGTTTCTTACATGAGT 881  
QY 221 GlyGlyLysAlaIleGlnHisLeuAlaLysGlnIleAsnArgPheHisPheAspIleLys 240  
DB 882 GGTGGAAGAGTATAGAACATTTGGCCAAAGACGGAATATTCATTTACATCAT 941  
QY 241 ProProLeuHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
DB 942 CCACCTATGCAAGATGCTAAGATTCGATTTTCTTTACGGGACCTTCACATATTTACT 1001  
QY 261 AspLysIleIleMetLysGlyGluGluGluGluGluGluGluGluGluGluGluGlu 280  
DB 1002 GATTAAGCTTAATACACACAGAGAAAGAGAGGACATTCAGAAAGGCGCAATCTGTCA 1061  
QY 281 SerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300  
DB 1062 TCAGCTCAGACATTTGCTGCTGCGGTACAGATCAGACAGGCTGCCACTTGCAGAAAAGA 1121  
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGlnAsnAsnAlaVal 320  
DB 1122 ACACATCGTGCTATTTGTTTTCAGACAGAAATTTGCTATCTCCAGCTTAAAGCAAGTA 1181  
QY 321 LeuValAlaSerGlyValAlaAlaSerAsnPheIleIleArgAlaIleuGlnIleLeu 340  
DB 1182 TTAGTTTATCTGAGAGGTGTGGCAAGTAACTTGATCCAGAAAGCATTTGAAATGTG 1241  
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360  
DB 1242 GCATATGCAACGACGTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1301  
QY 361 IleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
DB 1302 ATCATGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1361  
QY 381 IleGluGlyIleArgGlyArgProLysCysProLeuGlyValAspIleSerLysGluVal 400  
DB 1362 GTAGAAACATCCGATATGAAACCAAAATGTCCTTTGAAATAGACATATTCAGAGAGT 1421  
QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414  
DB 1422 GCAGAGCTGCCATTAAGTACCGCATTTAAATAATGGCACTT 1463

RESULT 10

AR541929 1416 bp DNA linear PAT 08-OCT-2004

DEFINITION Sequence 177 from patent US 6743619.

ACCESSION AR541929

VERSION AR541929.1 GI:53934009

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1416)

AUTHORS Tang, Y.T., Zhou, P., Goodrich, R., Liu, C., Asundi, V., Ren, F., Zhang, J., Zhao, Q.A., Yang, Y., Xue, A.J., Wehrman, T., Wang, J.-R., Wang, D. and Drenth, R.T.

TITLE Nucleic acids and polypeptides

JOURNAL Patent: US 6743619-A 177 01-JUN-2004;

FEATURES

location/Qualifiers

1..1416

/organism="unknown"

/mol\_type="genomic DNA"



## ORIGIN

## Alignment Scores:

|                        |          |               |      |
|------------------------|----------|---------------|------|
| Pred. No.:             | 5_9e-139 | Length:       | 1416 |
| Score:                 | 1747.00  | Matches:      | 340  |
| Percent Similarity:    | 100.00%  | Conservative: | 1    |
| Best Local Similarity: | 99.71%   | Mismatches:   | 0    |
| Query Match:           | 82.21%   | Indels:       | 0    |
| DB:                    | 6        | Gaps:         | 0    |

US-10-649-273-2 (1-414) x AR541929 (1-1416)

```

Qy 74 LysThrGlyValIleValProProAlaIleGlnIleuHisArgGluAsnIleGlnArg 93
Db 280 AGAAGAGGTGGATTGTTCTCCAGCAGCTCAACAGCTTCACAGAGAAATATTCACAGCA 339
Qy 94 IleValGlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAla 113
Db 340 ATAGTACAGAGACCTCTTCTGCGCAGTGAGTCTCCAGTGAACCTCTCAGCAATTGCA 399
Qy 114 ThrThrIleLeuSerProGlyLeuAlaLeuSerIleuGlyValGlyLeuSerPheSerLeuGln 133
Db 400 ACTACCAATMAACGAGACTTGCTTTAAGCCTGGAGTGGGCTTATCATTTAGCTTACAG 459
Qy 134 LeuValGlyIleuLeuValysProPheIleProIleHisIleMetGluAlaHisAlaLeu 153
Db 460 CTGGTAGCAGATTAAAGCCATTTCATTCCTCATTCATTCATTCAGAGCTCATGCACTT 519
Qy 154 ThrIleArgIleuThrAsnIysValGluPheProPheLeuValLeuLeuIleSerGly 173
Db 520 ACTATTAGTTGACCAATTAAGTGAATTCCTTTTATGATTTCTTGAATTTCTGAGAGT 579
Qy 174 HisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuLeuGlyLysSerIleu 193
Db 580 CACTGCTGTGGCATTAAGTTCAGAGAGTTCAGATTTTCGCTTTCGAAAGCTTTTG 639
Qy 194 AspIleAlaProGlyAspMetLeuAspIleValAlaArgArgLeuSerIleuIleHis 213
Db 640 GACATTCACCAAGTGCATGCTTGCACAGGTGGCAGAGACCTTTCTTAATAAATCAT 699
Qy 214 ProGluCysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaValGlnGlyAsn 233
Db 700 CCAAGAGTCCCAACCATGAGTGGGGAAGCCATPAGAACTTTGGCCAAACAGGAAT 759
Qy 234 ArgPheHisPheAspIleLysProProLeuHisIleSalAlaLysArgAspPheSerPhe 253
Db 760 AGATTTTCATTTGACATCAACCTCCCTTCATCATGCTTAATAATTCGATTTTCTTTT 819
Qy 254 ThrGlyLeuGlnHisValThrAspIleIleIleMetLysGluLysGluGluGlyIle 273
Db 820 ACTGGACTTAAACCGTTACTGATAAATATTAATGAAAAAGAAAAAGAGAGTAT 879
Qy 274 GluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMet 293
Db 880 GAGAGAGGGCAATTCCTGCTTCACAGCAGACATTCGTCACAGTACAGACACATG 939
Qy 294 AlaCysHisLeuValIysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeu 313
Db 940 GCATGTCATCTTGGAAGAAAGAACATCGGGCTATTGTTTGTGAACAGAGACTTG 999
Qy 314 LeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaAlaSerAspPheTyrlle 333
Db 1000 TTACCTCAAAATATATGAGTACTGCTGTCATCTGCTGTCGCAAGTACTTCATATTC 1059
Qy 334 ArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProPro 353
Db 1060 CGAGAGACTCTGGAATTTTAACCAACGACACAGTCACTTGTGTGTCTCTCC 1119
Qy 354 ArgLeuIleCysThrAsnAsnGlyValMetIleAlaIleArgGlnGlyIleGluArgLeuArg 373
Db 1120 AGACTATGCACTGATTAATGCAATTATGATTCATGAGATGATTAAGTAAAGACTACGTCT 1179
Qy 374 GlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrgluProLysCysProLeuGly 393

```

```

Db 1180 GCGTTGGGCAATTTACATGACATGAGCAGATCCGCTATGAAACCAAAATGCTCTTGGA 1239
Qy 394 ValAspIleSerLysGluValGlyValAlaSerIleLysValProGlnLeuLysMetGlu 413
Db 1240 GTAGCATATCATTAAGAAAGAGTTGGAGAACCTTCATTAAGTACCAATTAATAATGAG 1299
Qy 414 Ile 414
Db 1300 ATA 1302

RESULT 11
BC078974 1546 bp mRNA linear ROD 03-AUG-2004
LOCUS Rattus norvegicus cDNA clone IMAGE:7111906, partial cds.
DEFINITION BC078974
ACCESSION BC078974
VERSION BC078974.1 GI:50925879
KEYWORDS
SOURCE
ORGANISM Rattus norvegicus (Norway rat)
Rattus norvegicus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1546)
Strausberg,R.L., Fellings,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Sherman,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldi,M.F., Casavant,T.L.,
Schneitz,T.E., Brownstein,M.J., Uedlin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Rana,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richard,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Huliyil,S.N.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Heltón,E., Kettelman,M., Madan,A., Rodríguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodríguez,A.C., Grimwood,J., Schultz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,J., Smalins,D.E.,
Scherer,A., Schein,J.E., Jones,S.J., and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 1546)
Director MGC Project.
Submitted (02-AUG-2004) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcgaps-remail.nih.gov
Tissue Procurement: Howard Jacobs
CDNA Library Preparation: Express Genomics
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www.shgc.stanford.edu
Contact: (Dickson, Mark) mcdpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/UMN, at: http://image.llnl.gov
Series: IMAX Plate: 182 Row: F Column: 6
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein.
Location/Qualifiers
FEATURES

```



source

1. 1546  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10116"  
/clone="IMAGE:7111906"  
/cissue\_type="Testis, rat (Brown Norway)"  
/clone\_id="NH\_MGC\_237"  
/lab\_host="DH108"  
/note="Vector: pExpress1"  
144. .>1546  
/codon\_start=1  
/product="Unknown (protein for IMAGE:7111906)"  
/protein\_id="AAH78974.1"  
/db\_xref="GI:50926880"  
/translation="MMLSTKAGAPRRPRNVRGPIRRFNVPRALEHHKLVGIEIT  
SCDDPAAVDENGVNIGEAHSOTFVHLKTGIVPVPAQLRENTORIVEALSIS  
GVSPDLASAIATTIKRGALSLGVGSFQVUNQKRPPIPHHEAAHLLTRLTTHK  
VGPFLVLISGHCLLAVQSVDFLLKSLDIAFGMDLVARSLIKPEEST  
MSGKAIEMIAKGNRFFHTINPMQNAKNCDSFTGLQHTVDLITHKKEBGIENG  
QLISSADIAAAVQATACHLAKRTHRALIFCOQKMLSPANNVLVSGVASLTYR  
RALEIVANATQCTLLCPRLCTDNGIMIMWNGIERLRAGILIHVEDIRYRKAGS  
IYSMLRVLATLETDLSVSTHTVAHSPLNSSRGANIQTSMSCITQTYMRTYRHL  
NINERSKKKKKKKK"

ORIGIN

Alignment Scores:

| Pred. No.:             | 4.9e-137 | length:       | 1546 |
|------------------------|----------|---------------|------|
| Score:                 | 1725.00  | Matches:      | 335  |
| Percent Similarity:    | 91.00%   | Conservative: | 19   |
| Best Local Similarity: | 86.12%   | Mismatches:   | 35   |
| Query Match:           | 81.18%   | Indels:       | 0    |
| DB:                    | 10       | Gaps:         | 0    |

US-10-649-273-2 (1-414) x BC078974 (1-1546)

QY 1 MetLeuIleuThrsThraIagIValPhePheIysProSerIysArgIysValTyr 20  
DB 144 ATGCTATGTTAGTAAGACAGACAGAGCTATTCGACACACCAAGAGTAAGTTCGT 203  
QY 21 GluPheLeuArgSerPheLeuPheHisProGlyThrLeuPheLeuHisIleValLeu 40  
DB 204 GGAATTATTAAGATTTTAACTTCACTTCAAGACACTTTTTCATCTTAACTGCTG 263  
QY 41 GlyIleGluThrsCysAspAspThrAlaAlaValAlaValAspGluThrGlyAsnVal 60  
DB 264 GGAATTGAACCGACTGTGATGACACAGACGCTGCTGTGATGATGAAGAAAGTG 323  
QY 61 LeuGlyValAlaIleHisSerGlnThrGluValHisIleuIysThrGlyIleValPro 80  
DB 324 CTGGGAGAGCACTGCCTCCAGACGGAAGTCCATCTGAACAGGTGGGATTGTTCT 383  
QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
DB 384 CCAGTGCCTCAACAGCTTCAACAGAAATATTCAGCAATATAGAGAAAGCTCTTCT 443  
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleIysProGlyLeu 120  
DB 444 GCCAGTGGGCTCTCCCAACGACCTCTCAGCAATTCACATCACTCAACACAGACTG 503  
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerIleuGlnLeuValGlyIleuLeuIys 140  
DB 504 GCCCTAAGCTGGGGGTGGCTTATCTTATAGCGTACAGCTAGTAATCACTTTAAAG 563  
QY 141 ProPheIleProIleHisIsmetGluAlaHisIleAlaLeuThrIleArgLeuThrAsnIys 160  
DB 564 CCATTATCCCATTCATCATGAGAGCTCAGCGCTGACTATATAGCTGACCAACAAA 623  
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
DB 624 GTCGATTCCTTTTATGATTTCTTGAATTCGAGAGCCACTGCTGGCCCTTATGTT 683  
QY 181 GlnGlyValSerAspPheLeuLeuGlyIysSerIleuAspIleAlaProGlyAspMet 200

DB 684 CAGAGTGTTCAGATTTTCTGCTCTCGGGAAGTCCCTGACATAGCCCGCAGCATG 743  
QY 201 LeuAspIysValAlaArgArgLeuSerIleuIleIysHisProGluCysSerThrMetSer 220  
DB 744 CTTCACAGAGTGGCAGAGAGCTTCTTTATCAATCAATCCAGATGTTTACATAGT 803  
QY 221 GlyIleIysAlaIleGluHisIleuAlaIysGlnGlyAsnArgPheHisPheAspIleIys 240  
DB 804 GTGGGAAGCTATAGAACATTTGGCCAAAGAAATAGATTCACCTTATCATCAAT 863  
QY 241 ProProLeuHisHisAlaIysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
DB 864 CCACCCATGACAGATGCTTAAACCTGATTTTCTTTACGGGACTTCAACATGTCAC 923  
QY 261 AspIysIleIleMetIysGlyGluIysGluGlyIleGlyIysGlyGlnIleLeuSer 280  
DB 924 GATTAAGCTAATTAACACACAGAAAGAAAGAGAGCATTTGAAAGGGCAATCTGTCA 983  
QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisIleuValIysArg 300  
DB 984 TCAGCCGACAGCATTTGCTGCTGCGGTACAGCAGCAACAGGCTGCCACTTGCAGAAA 1043  
QY 301 ThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320  
DB 1044 ACACATGCTGCTATCTGTTGTTGACAGCAAAAATTTGCTATCTCCAGCTTAAGCAGTA 1103  
QY 321 LeuValAlaSerGlyValAlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeu 340  
DB 1104 TTATGTTGCTGAGAGGTGTGACAGTACTTGTACATCCAGAGCATTTGAAATTTGTA 1163  
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360  
DB 1164 GCAATGCAACACATGACTTGTGTGCTCCCTCCGAGACTGTGACATGACATGCT 1223  
QY 361 IleMetIleAlaIlePasnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
DB 1224 ATCATGATTCATGATGAAATGGAATTAAGATGCTGCTGCGCATTTTACATGAT 1283  
QY 381 IleGluGlyIleArgIleArgIleProIys 389  
DB 1284 GTAGAGACATCCGATAGCAACAAAG 1310

RESULT 12

AR428809 1526 bp DNA linear PAT 18-DEC-2003

DEFINITION Sequence 23 from patent US 6642041.

ACCESSION AR428809

VERSION AR428809.1 GI:40188595

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1526)

AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.

TITLE Polynucleotides encoding a novel metalloprotease, MP-1

JOURNLS Patent: US 6642041-A 23 04-NOV-2003;

FEATURES

Location/Qualifiers

1. 1526

source

1. 1526

/organism="unknown"

/mol\_type="genomic DNA"

ORIGIN

Alignment Scores:

| Pred. No.:             | 3.95e-108                                 | length:       | 1526 |
|------------------------|---|---------------|------|
| Score:                 | 1385.00 <td>Matches:</td> <td>267</td>    | Matches:      | 267  |
| Percent Similarity:    | 100.00% <td>Conservative:</td> <td>0</td> | Conservative: | 0    |
| Best Local Similarity: | 100.00% <td>Mismatches:</td> <td>0</td>   | Mismatches:   | 0    |
| Query Match:           | 65.18% <td>Indels:</td> <td>0</td>        | Indels:       | 0    |
| DB:                    | 6 <td>Gaps:</td> <td>0</td>               | Gaps:         | 0    |

US-10-649-273-2 (1-414) x AR428809 (1-1526)

QY 148 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIysValGluPheProPheLeuVal 167

```

Db      1 ATGGAGGCTCATGCACTTACTATTAGGTTACCAATAAGTAGAATTTCTTTTATAGTT
Qy      166 LeuLeuLeuSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu
Db      61 CTTTTCATTTCTTGAGAGTCACTGTCTGTGGCATTTAGTTCAGAGGATTTTCAGATTTTCTG
Qy      188 LeuLeuGlyLeuSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArg
Db      121 CTTCTTGGAAGCTTTTGACATAGACCAAGTGACATGCTTGACAGAGTGGCAAGAGA
Qy      208 LeuSerLeuIleLeuHisSProGluCysSerThreSerGlyGlyValAlaIleGlnHis
Db      181 CTTTCTTTATATAAACAATCCAGAGTGTCTCCACCAATGAGTGGGAAGCCATAGAACAT
Qy      228 LeuAlaLeuGlnGlyAsnArgPheHisPheAspIleLeuProPheLeuHisLeuAlaLeu
Db      241 TTGGCCAAACAGAAATATATTTCAATTTGACATCAACCTCCCTTGACATCATCTTAA
Qy      248 AsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspIleIleMetLeuLeu
Db      301 AATTGATTTTCTTTTACTGATCTTCAACAGCTTATGATTAATAATATATATAAAG
Qy      268 GlyLeuGlnGlyIleGlyLeuGlnIleLeuSerSerAlaAlaAspIleAlaAla
Db      361 GAAAGAGAGAGATTTGAGAGGAGGCAATCTGTCTTACAGACAGACATTTGCTGCC
Qy      288 ThrValGlnHisThrMetAlaCysHisLeuValValArgThrHisAlaGlnIleLeuPhe
Db      421 ACGTACACACACCAATGGCATGTCATCTGTGTGAAGAACACATCGGCTATTTCTGTTT
Qy      308 CysLeuGlnArgAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGlyVal
Db      481 TGTACAGAGAGATTTGTTTACTTCATTAATATATGCACTGCTGTGCACTGTGCTGTC
Qy      328 AlaSerAspPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr
Db      541 GCAAGTACTCTCATATCCGACAGAGCTCTGGAATTTTAAACAACGACACAGCTGCACT
Qy      348 LeuLeuCysProProAlaGlyLeuCysThrAspAsnGlyIleMetIleAlaTTPAsnGly
Db      601 TTGTTGTCTCTCTCCCAACATATGCACTGATATAGTGCATTTATGATGCAATGCAATGCT
Qy      368 IleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlyGlyIleArgTyrGln
Db      661 ATTGAAGAGCTACGTGCTGCTGGCATTTTACATATCATAAGAGCATCTCCGTATGAA
Qy      388 ProLeuCysProLeuGlyValAspIleSerLeuGlyValGlyGlnAlaSerIleLeuVal
Db      721 CCAAAATGTCCTTTGGAGTAGACATATCAAAAGAAAGTTGGAGAACTTCCATTAAGTA
Qy      408 ProGlnLeuLeuMetGluIle 414
Db      781 CCACAAATTAATAATGAGATA 801

```

RESULT 13  
 BX934991 1522 bp mRNA linear VRT 02-FEB-2004  
 LOCUS Gallus gallus finished cDNA, clone CHEST189114.  
 DEFINITION BX934991  
 ACCESSION BX934991.1 GI:41635519  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Gallus gallus (chicken)  
 Gallus gallus  
 Bkaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 1 (bases 1 to 1522)  
 Boardman, P. B., Bonfield, J. K., Brown, W. R. A., Carder, C., Chalk, S. E.,  
 Croning, M. D. R., Davies, R. M., Francis, M. D., Grafham, D. V.,  
 Hubbard, S. J., Humphray, S. J., Hunt, P. J., Maddison, M., McLaren, S. R.,  
 Nibbelc, D., Overton, I. M., Rogers, J., Scott, C. E., Taylor, R. G.,  
 Tickle, C. and Wilson, S. A.

TITLE Direct Submission  
 JOURNAL Submitted (02-FEB-2004) Sanger Institute, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: chickest@hms.umbc.ac.uk  
 COMMENT BSRRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA  
 sequencing project.  
 This sequence is from the  
 BSRRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,  
 from a library constructed by Elizabeth Bosch. cDNA was prepared  
 from RNA extracted from muscle, normalised, and poly A-tailed.  
 EcoRI-NotI cut cDNA was then ligated into the vector. Vector:  
 pBlueScript II KS(+); Site\_1: EcoRI; Site\_2: NotI Host: Escherichia  
 coli DH10B.

FEATURES  
 source  
 1. 1522  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /script="layer and broiler"  
 /db\_xref="taxon:9031"  
 /clone="CHEST189114"  
 /clone\_1db="CEORBN11"  
 /dev\_stage="adult"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 3,56e-106 Length: 1522  
 Score: 1362.00 Matches: 258  
 Percent Similarity: 83.20% Conservative: 54  
 Best Local Similarity: 68.80% Mismatches: 63  
 Query Match: 64.09% Indels: 0  
 DB: Gaps: 5

US-10-649-273-2 (1-414) x BX934991 (1-1522)  
 Qy 37 LysIleValIleuGlyIleGlnThrSerCysAspAspThrAlaAlaValAlaAspGlu 56  
 Db 254 AGACTTGTGCTGGGCATCGAAACAGCTGTACACAGAGCGCGCTGTCAGACAG 313  
 Qy 57 ThrGlyAsnValIleuGlyGlnAlaIleHisSerGlnThrGluValHisLeuLeuThrGly 76  
 Db 314 CGGGACAGGCTGCTGGAGAGAGCCCTGCAGAGCCAAAGAGGCTCACTCAAGACAGT 373  
 Qy 77 GlyIleValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGln 96  
 Db 374 GGATTAATTCCTCATATGACACAGAGCTTCAAGAGAGATTCACAGAGTAAAG 433  
 Qy 97 GluAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIle 116  
 Db 434 GAACCACTCACTGCAGAGGAGTTCTGTAATGAACTGCTGCTATTCAGCACTACAGTG 493  
 Qy 117 LysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnValGly 136  
 Db 494 AAACCAAGACTTGGCTGAGCTGAGAGTGGAGCTGACGTACAGCTTACAGCTGGTGAAC 553  
 Qy 137 GlnLeuLeuValSProPheIleProIleHisHisMetGlnAlaHisAlaLeuThrIleArg 156  
 Db 554 AGTACCAAGAGCTTTTCATATCCCATTCATATCAGAGGCTCAGAGCATTAACATGCA 613  
 Qy 157 LeuThrAsnValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeu 176  
 Db 614 CTGACAGAGCAAGTAAATTTCTTCTTACTTCTTACTCTCCGAGAGTCACTGATC 673  
 Qy 177 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyValSerLeuAspIleAla 196  
 Db 674 TTGGCAGTAGACACAGAGGTTTCAGATTTCTTCTGCTGACATTCATCATATATACCA 733  
 Qy 197 ProGlyAspMetLeuAspIleValAlaArgArgLeuSerLeuIleLeuHisProGluCys 216  
 Db 734 CCAAGTACATGCTGAGTAAAGTAGCAAGAGGCTCTTTTATGATAGACCCGAGATGC 793  
 Qy 217 SerThreSerGlyGlyValAlaIleGlnHisLeuAlaValGlnGlyValAspArgPheHis 236  
 Db 794 CACGCAATGCTGGGGAGAGGCAATAGACACCTGCTCAACCGAGAGCTGCAACAG 853

|            |  |      |  |                             |
|------------|--|------|--|-----------------------------|
| Qy         |  | 237  | Pheasp1lelysProPlelhis1sh1alalyasenCyAspPhSeSphenChilyeu         | 256                         |
| Db         |  | 854  | TACACTTCAGACTCCCATGCACACAGATGCTACTGTGATTCTTTCTTCCGGACTT          | 913                         |
| Qy         |  | 257  | Glnh1svAlThrAspLyslleIleMelysLysgLysgJugLylYleqlubysqLy          | 276                         |
| Db         |  | 914  | CAGAGCCTTGTCACAAAGCCATTCCTTCAGAAAAGAAAAAGAAAGATTCAGAAGGG         | 973                         |
| Qy         |  | 277  | Gln1leLeuSerSerAlAlAsp1leAla1thrValGlnh1sthrMetAlacyshis         | 296                         |
| Db         |  | 974  | GAAATTCCTGTCCTCGTTTAGACATCCCTGCTGTGACACAGCAAGTGAAGTCCTCAT        | 1033                        |
| Qy         |  | 297  | LeuValysArgThriAsrAla1leLeuPheCyLySGlnArgAspLeuProGln            | 316                         |
| Db         |  | 1034 | ATTATCCAGCGGACACACCAGCCATGCTCTTCATGAAAAAACAACATATTATTACCA        | 1099                        |
| Qy         |  | 317  | AsnAsnAlaValLeuValAlaSerLygLyValAlaSerAsnPhetyrlleArgArgAla      | 336                         |
| Db         |  | 1094 | AAACCTGCACCTGCTGTGTATCAGAGAGAGTTCAGATTAATCATATATACGAAAAGGA       | 1155                        |
| Qy         |  | 337  | Leuql1leLeuThrAsnAla1thrgLncysrh1leuLeuCySProProArgLeuCyS        | 356                         |
| Db         |  | 1154 | CTGAGACTCTGGCAAATGCAAAAGCGTTTGGCTTTTCGTCTCCCTCCACAGCGCTGCC       | 1213                        |
| Qy         |  | 357  | ThrsPasnnglyIleMet1leAlatrpAsnnglyIleGlnArgLeuArgAlaglyLeuengly  | 376                         |
| Db         |  | 1214 | ACCATAATAGTGTATTATGATTCATGAAATGCATGAAAGGTTCGTCGACAGATGTGCT       | 1277                        |
| Qy         |  | 377  | IleLeuH1saSp1leegLugLy1leArgTyrgLuproLyCySProlenglyValAsp1le     | 396                         |
| Db         |  | 1274 | ATTTTATTCAGTACTAGTATGATGCGATCCGCTACGAAACCAAAAGCTCCCTTGGAATTATATT | 1333                        |
| Qy         |  | 397  | SerLySGluValAlglyGluAlaSer1leLySVa1ProGlnLeuLys                  | 411                         |
| Db         |  | 1334 | TCCAAAGAGTGTGAAGAGATTCATCCAAGTGCACCAAGACTAAG                     | 1378                        |
| RESULT 14  |  |      |  |                             |
| LOCUS      | BC038910   |      | 1017 bp  | mRNA linear ROD 21-OCT-2002 |
| DEFINITION | Mus musculus O-sialylcoprotein endopeptidase-like 1, mRNA (cdna  |      |  |                             |
| ACCESSION  | BC038910   |      |  |                             |
| VERSION    | BC038910.1   |      | GI:24433548  |                             |
| KEYWORDS   |  |      |  |                             |
| SOURCE     | Mus musculus (house mouse)   |      |  |                             |
| ORGANISM   | Mus musculus   |      |  |                             |
| REFERENCE  | Eukarya; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;<br>Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.<br>1 (bases 1 to 1017)   |      |  |                             |
| AUTHORS    | Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,<br>Klausner, R.D., Collins, F.S., Wagner, L., Shemmen, C.M., Schuler, G.D.,<br>Altshuler, S.F., Zeeberg, B., Butow, K.H., Schaefer, C.F., Bat, N.K.,<br>Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hasteh, P.,<br>Datchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,<br>Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.M.,<br>Schneerz, T.E., Brownstein, M.J., Uesdin, T.B., Toshiyuki, S.,<br>Carancini, P., Prange, C., Rada, S.S., Loquellano, N.A., Peters, G.J.,<br>Abramson, R.D., Mullahy, S.U., Bosak, S.A., McEwan, P.J.,<br>McNernan, K.U., Malek, J.A., Gunaratne, P.H., Richards, S.,<br>Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Holly, S.W.,<br>Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,<br>Fahey, J., Helton, E., Kettman, M., Madan, A., Young, A.C., Rodrigues, Y.,<br>Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,<br>Bonifard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,<br>Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,<br>Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smilins, D.E.,<br>Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A. |      |  |                             |
| TITLE      | Generation and initial analysis of more than 15,000 full-length<br>human and mouse cDNA sequences  |      |  |                             |
| JOURNAL    | Proc. Natl. Acad. Sci. U.S.A.  |      | 99 (26), 16899-16903   | (2002)                      |
| MEDLINE    | 22388257   |      |  |                             |
| PUBMED     | 12477932   |      |  |                             |
| REFERENCE  | 2 (bases 1 to 1017)  |      |  |                             |

**AUTHORS**  
Strausberg, R.  
**TITLE**  
Direct Submission  
**JOURNAL**  
Submitted (25-OCT-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

**REMARK**  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
**COMMENT**  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
[info@bcsc.bc.ca](mailto:info@bcsc.bc.ca)  
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Len Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

**FEATURES**  
source  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRAK Plate: 86 Row: F Column: 12.  
Location/Qualifiers  
1..1017  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5053559"  
/tissue\_type="Liver, normal. 5 month old male mouse."  
/clone\_id="NCI CGAP\_L19"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"

**ORIGIN**  
.  
.

**Alignment Scores:**  
Pred. No.: 7,91e-99 Length: 1017  
Score: 1273.00 Matches: 241  
Percent Similarity: 92.73% Conservative: 14  
Best Local Similarity: 87.64% Mismatches: 20  
Query Match: 59.91% Indels: 0  
Gaps: 0

US-10-649-273-2 (1-414) x BC038910 (1-1017)

Dy 140 LysPProhelelePcoIlehiShiSMeGuaIaShiSaIaleuThrIleArgLeuThraSn 159

Dd 2 AAGCATTCATCCCGATTCATCATCAGAGGCTCAAGCAGTCACTATTAGGCTCACCAAT 61

Dy 160 LysValGluPhePProPheLeuValleuLeuIleSerGlyGlyHisGlyLeuAlaLeu 179

Dd 62 AAAGTAGAATTCCTTTTATGATTCCTTTGATTTCTGGGGCTACCTGTTGGCATTA 121

Dy 180 ValGlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAsp 199

Dd 122 GTCCAGAGTGTTCGAGATTTCTGCTCTCTTGGAAGTCTTTGAGACATAGCACAGGGAC 181

Dy 200 MetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMet 219

Dd 182 ATGCTTGCACAGGTGGCAAGAAGATTTCTTAAATCAACATCCAGAAATGTTCTACATG 241

Dy 220 SerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIle 239

Dd 242 AGTGGTGAAAAGCTATAGAACAGTGGCCAAAGACGGAATAAGATTTCATTACTATC 301

Dy 240 LysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisVal 259

Dd 302 AATCCACTATGACAGAAATGCTAAGAAATTCGATTTTCTTTCACGGGACTTCAACATATT 361

|            |   |   |                             |
|------------|---|---|-----------------------------|
| Qy         | 260   | ThrsApysIeIleMeTyLysGluYsgLuglylIleGluYsgLglnIleu               | 279                         |
| Db         | 362   | ACTGATTAAGCTAAATTAACACACAGGAAAGAAAGAGCAATTGAGAGGGCAATCTG        | 421                         |
| Qy         | 280   | SeSeSerAlaAlaAspIleAlaIleAlaIleValGlnIsthrMetAlaCyshIleuValIys  | 299                         |
| Db         | 422   | TCATCAGGTGACGACATTCGTCTGCGCGGTACAGCATGCACAGCGGTCCACTTGCAGAA     | 481                         |
| Qy         | 300   | ArgThrHisAlaGAlaIleLeuPheCysYlsgLlnArgAspLeuLeuProGlnAsnAsnAla  | 319                         |
| Db         | 482   | AGAACACATCCCGCATATCTCTTTTGGACAGAAAGAAATTTGGTCTCTCCAGCTAACGCA    | 541                         |
| Qy         | 320   | ValIeuValAlaAspGlyGlyValAlaIleSerAsnPheTyrIleArgArgAlaIleGluIle | 339                         |
| Db         | 542   | GTATTAGTGTATCTGAGGAGTGTGCAGAGTAACCTGTACATCCGAAAGCAATGGAAAT      | 601                         |
| Qy         | 340   | LeuThrAsnAlaIleThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsn    | 359                         |
| Db         | 602   | GTCCGAATGCACCGCAGTGCACGTTGTGTGTCTCACCCTCCAGACTGTGCATGCACAT      | 661                         |
| Qy         | 360   | GlyIleMetCileAlaIleAspAsnGlyIleGluArgLeuArgAlaGlyIleuGlyIleuHis | 379                         |
| Db         | 662   | GGCATCATGATTCATGCATGGAATGGAATTAAGATTACGTGCGCTTGCGGCTTTATCAT     | 721                         |
| Qy         | 380   | AspIleGlyGlyIleArgTyrGlyProLysCysProLeuGlyValAspIleSerLysGlu    | 399                         |
| Db         | 722   | GATGAGAGAACATCCGATATGAAACCAAAATGCTCCTTGGAGTGAATATCCAGAGAA       | 781                         |
| Qy         | 400   | ValGlyGluAlaSerIleIleYsValProGlnIleuLysMetGluIle                | 414                         |
| Db         | 782   | GTTCGAGAGCTGCCATTAAGTACCGCATTTAAATAATGGACATT                    | 826                         |
| RESULT 15  |   |   |                             |
| LOCUS      | EX330963  | 1558 bp   | mRNA linear VRT 30-MAR-2004 |
| DEFINITION | Gallus gallus finished cDNA, clone CHEST62n16.                          |   |                             |
| ACCESSION  | EX330963  |   |                             |
| VERSION    | EX330963.2  | GI:46016890   |                             |
| KEYWORDS   |   |   |                             |
| SOURCE     |   |   |                             |
| ORGANISM   | Gallus gallus (chicken)   |   |                             |
| REFERENCE  |   |   |                             |
| AUTHORS    | 1 (bases 1 to 1558)   |   |                             |
|            | Boardman, P.E., Bonfield, J.K., Brown, W.R.A., Carder, C., Chalk, S.E., |   |                             |
|            | Croning, M.D.R., Davies, R.M., Francis, M.D., Graham, D.V.,             |   |                             |
|            | Hubbard, S.J., Humphray, S.U., Hunt, P.J., Maddison, M., McLaren, S.R., |   |                             |
|            | Niblett, D., Overton, I.M., Rogers, J., Scott, C.E., Taylor, R.G.,      |   |                             |
|            | Tickle, C. and Wilson, S.A.   |   |                             |
| TITLE      | Direct Submission   |   |                             |
| JOURNAL    | Submitted (29-MAR-2004) Sanger Institute, Hinxton, Cambridgeshire,      |   |                             |
|            | CB10 1SA, UK. E-mail enquiries: chickens@bms.unist.ac.uk                |   |                             |
| COMMENT    | On Apr 1, 2004 this sequence version replaced gi:41631491.              |   |                             |
|            | BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA       |   |                             |
|            | sequencing project.   |   |                             |
|            | This sequence is from the   |   |                             |
|            | BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,         |   |                             |
|            | from a library constructed by Elizabeth Bosch. cDNA was prepared        |   |                             |
|            | from RNA extracted from limbs,  |   |                             |
|            | and poly A-titrmed. EcoRI-NotI cut cDNA was then ligated into the       |   |                             |
|            | vector. Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI       |   |                             |
|            | Host: Escherichia coli DH10B.   |   |                             |
|            | Location/Qualifiers   |   |                             |
|            | 1..1558   |   |                             |
| FEATURES   |   |   |                             |
| Source     | /organism="Gallus gallus"   |   |                             |
|            | /mol_type="mRNA"  |   |                             |
|            | /strain="White Leghorn, Hisex"  |   |                             |
|            | /db_xref="taxon:9031"   |   |                             |
|            | /clone="CHEST62n16"   |   |                             |
|            | /clone_lib="CSEQC120"   |   |                             |
|            | /dev_stage="stage 36"   |   |                             |

[illegible]

Db 1386 GGATGCTGTAATTTATACAGTACTGATGGCATCGCTATGACCAAGCTCCCTTGA 1445  
 Oy 394 ValAspIleSerIysGluValGlyGluAlaSerIleIysValProGlnLeuLys 411  
 Db 1446 ATTGATATTTCCAAAAGAGTTGAGAGGATTCATCCAGAGTGCACGACTAAGG 1499

Search completed: June 16, 2005, 23:18:30  
 Job time : 4983.35 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: June 16, 2005, 12:03:59 ; Search time 598.5 Seconds  
(without alignments)  
4094.859 Million cell updates/sec

Title: US-10-649-273-2  
Perfect score: 2125  
Sequence: 1 MLITKTAGVFPKPSKRYK...DISKEVGEASIKVPOLMEI 414

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 segs, 295870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seg length: 0  
Maximum DB seg length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODE=frame+.p2n.model -DEV=xlp  
-Q=/cgn2.1/USPTO.spool.p/US10649273/runat.15062005.11415.6022/app.query.fasta\_1.1429  
-DB=N Geneseg.16Dec04 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blonsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10649273.cgn.1.1.1063 @runat.15062005.11415.6022 -NCPU=6 -ICPU=3  
-NO NMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N\_Geneseg.16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 2125  | 100.0       | 2058   | 6     | ABA93268 Human O-s |
| 2          | 2125  | 100.0       | 2197   | 6     | AB876635 DNA encod |
| 3          | 2125  | 100.0       | 2572   | 8     | ABT23207 Human pro |
| 4          | 2088  | 98.3        | 1820   | 6     | AAd46856 Human g1y |
| 5          | 2088  | 98.3        | 1820   | 8     | ACA60887 Human cDN |

|    |        |      |        |    |          |                    |
|----|--------|------|--------|----|----------|--------------------|
| 6  | 2088   | 98.3 | 1821   | 10 | AB857020 | AB857020 cDNA enco |
| 7  | 1944   | 91.5 | 2208   | 10 | ADA52832 | Ada52832 Human cod |
| 8  | 1944   | 91.5 | 2890   | 12 | ADQ24627 | Adq24627 Human sof |
| 9  | 1747   | 82.2 | 1416   | 8  | ABX70950 | Abx70950 Novel hum |
| 10 | 1549   | 72.9 | 1572   | 6  | ABQ75508 | Abq75508 Murine si |
| 11 | 1385   | 65.2 | 1526   | 6  | AB876639 | Ab876639 DNA encod |
| 12 | 1001.5 | 47.1 | 3358   | 10 | ADB31345 | AdB31345 Human dia |
| 13 | 714.5  | 33.6 | 1601   | 4  | ABL24633 | AbI24633 Drosophi1 |
| 14 | 700.5  | 33.0 | 1385   | 4  | AAH15110 | AAH15110 Human cDN |
| 15 | 687.5  | 32.4 | 3656   | 4  | ABL24632 | AbI24632 Drosophi1 |
| 16 | 662.5  | 31.2 | 1557   | 3  | AAC38454 | AAC38454 Arabidops |
| 17 | 629    | 29.6 | 571    | 4  | AAH12983 | AAH12983 Human cDN |
| 18 | 599    | 28.2 | 2734   | 5  | AA884622 | AA884622 DNA encod |
| 19 | 531.5  | 25.0 | 4360   | 6  | AA48239  | AA48239 Ehrlichia  |
| 20 | 524    | 24.7 | 1146   | 8  | ACA26804 | ACA26804 Prokaryot |
| 21 | 502    | 23.6 | 1044   | 8  | ACA39102 | ACA39102 Prokaryot |
| 22 | 502    | 23.6 | 94750  | 4  | AAF28551 | AAf28551 Genomic f |
| 23 | 501    | 23.6 | 1053   | 12 | ADL03120 | ADL03120 DNA encod |
| 24 | 500    | 23.5 | 1092   | 6  | ABQ90383 | Abq90383 M. capuli |
| 25 | 499.5  | 23.5 | 1206   | 11 | ABD02197 | ABd02197 Pseudomon |
| 26 | 494.5  | 23.3 | 1032   | 8  | ACA43173 | ACA43173 Prokaryot |
| 27 | 492.5  | 23.2 | 1026   | 4  | AA554064 | AA554064 Pseudomon |
| 28 | 492.5  | 23.2 | 1026   | 8  | ACA42146 | ACA42146 Prokaryot |
| 29 | 492.5  | 23.2 | 1026   | 10 | ADG73341 | Adg73341 P aerugin |
| 30 | 492.5  | 23.2 | 1059   | 11 | ABD02280 | ABd02280 Pseudomon |
| 31 | 490.5  | 23.1 | 1026   | 10 | ADG73343 | Adg73343 P aerugin |
| 32 | 484.5  | 22.8 | 1014   | 4  | AA552570 | AA552570 E. coli D |
| 33 | 484.5  | 22.8 | 1014   | 8  | ACA51431 | ACA51431 Prokaryot |
| 34 | 484.5  | 22.8 | 1014   | 8  | ACA32641 | ACA32641 Prokaryot |
| 35 | 483.5  | 22.8 | 1014   | 8  | ACA54107 | ACA54107 Prokaryot |
| 36 | 482.5  | 22.7 | 1020   | 8  | ACA44384 | ACA44384 Prokaryot |
| 37 | 482.5  | 22.7 | 1074   | 10 | ADF02056 | ADf02056 Bacteriat |
| 38 | 481.5  | 22.7 | 1014   | 4  | AA556045 | AA556045 Salmonell |
| 39 | 479    | 22.5 | 1315   | 2  | AAQ27645 | AAq27645 Glycyprot |
| 40 | 475.5  | 22.4 | 1029   | 4  | AA553309 | AA553309 Haemophil |
| 41 | 475.5  | 22.4 | 1029   | 8  | ACA34150 | ACA34150 Prokaryot |
| 42 | 475.5  | 22.4 | 110000 | 2  | AAT42063 | Continuation (6 of |
| 43 | 474.5  | 22.3 | 1032   | 11 | ACH96255 | ACH96255 Klebsiell |
| 44 | 472.5  | 22.2 | 1014   | 10 | ACF71364 | ACf71364 Photorhab |
| 45 | 472.5  | 22.2 | 110000 | 10 | ACF67367 | Continuation (46 o |

## ALIGNMENTS

|          |  |
|----------|--|
| RESULT 1 |  |
| ABA93268 | standard; cDNA; 2058 BP.   |
| ID       | ABA93268;  |
| AC       | ABA93268;  |
| XX       |  |
| DT       | 19-APR-2002 (first entry)  |
| XX       |  |
| DE       | Human O-sialoglycoproteinase-like protein encoding cDNA SEQ ID NO:1. |
| XX       |  |
| KW       | Human; O-sialoglycoproteinase-like protein; OSGLP; enzyme; gene; ss. |
| XX       |  |
| OS       | Homo sapiens.  |
| XX       |  |
| FH       | Key  |
| FT       | CDS  |
| FT       | Location/Qualifiers  |
| FT       | /*tag= a   |
| FT       | /product= "O-sialoglycoproteinase-like protein"                      |
| PN       | CN118550-A.  |
| XX       |  |
| PD       | 24-OCT-2001.   |
| XX       |  |
| PF       | 19-APR-2000; 2000CN-00106834.  |
| XX       |  |
| PR       | 19-APR-2000; 2000CN-00106834.  |
| XX       |  |
| PA       | (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.                          |
| XX       |  |

|   |   |                 |
|---|---|-----------------|
| PI  | Mao Y, Xie Y;   |                 |
| XX  |   |                 |
| DR  | WPI: 2002-115090/16.  |                 |
| DR  | P-ESDB; ABB05481.   |                 |
| XX  |   |                 |
| PT  | O-sialoglycoproteinase-like protein and encoding polynucleotide, useful   |                 |
| XX  | for diagnosing, preventing and treating related diseases.                 |                 |
| PS  | Claim 5; Page 29-30 (Disclosure); 38pp: Chinese.                          |                 |
| XX  |   |                 |
| CC  | The present sequence encodes human O-sialoglycoproteinase-like protein    |                 |
| CC  | (OSGPLP). The present invention also describes: (1) the preparation of    |                 |
| CC  | the OSGPLP protein; (2) applying the OSGPLP protein in diagnosis; (3) the |                 |
| CC  | prevention and/or treatment of related diseases; (4) utilizing the OSGPLP |                 |
| CC  | protein in screening its agonist, excitomotor and inhibitor and preparing |                 |
| CC  | an antibody against the OSGPLP protein; and (5) the use of the OSGPLP     |                 |
| CC  | polynucleotide sequences, proteins, agonists, excitomotors, inhibitors    |                 |
| CC  | and antibodies in treating diseases related to the abnormal OSGPLP gene   |                 |
| CC  | and in preparing the medicine composite for the treatment                 |                 |
| XX  |   |                 |
| SQ  | Sequence 2058 BP; 637 A; 400 C; 410 G; 611 T; 0 U; 0 Other;               |                 |
| Alignment Scores:                           |   |                 |
| Pred. No.:                                  | 1,94e-212   | Length: 2058    |
| Score:                                      | 2125.00   | Matches: 414    |
| Percent Similarity:                         | 100.00%   | Conservative: 0 |
| Best Local Similarity:                      | 100.00%   | Mismatches: 0   |
| Query Match:                                | 100.00%   | Indels: 0       |
| DB:   | 6   | Gaps: 0         |
| US-10-649-273-2 (1-414) x ABA93268 (1-2058) |   |                 |
| OY  | 1 MetLeuIleLeuThrIysThrAlaGlyValPhePheLysProSerIysArgLysValTyr 20         |                 |
| Db  | 110 ATGTAATTCCTTGACATACACTGACGAGAGGTTTTTTTAAACCATCAAAAAGAAAGTTAT 169      |                 |
| OY  | 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40        |                 |
| Db  | 170 GAATTTTAAAGAAAGTTTAAATTTTCACTCTGGAACACTATTCTTCATATAAAATGAAATG 229     |                 |
| OY  | 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60           |                 |
| Db  | 230 GGAATTTGAAACTAGTTGTGATGATACAGCAGCTGCTGTGTGATGATGAAATCTGAAATGTG 289    |                 |
| OY  | 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80           |                 |
| Db  | 290 TTGGGAGAAACAATACATCCCAAACTGAAGTCAATTTAAAAACAGGTGGATTTGCT 349          |                 |
| OY  | 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100       |                 |
| Db  | 350 CCAGACAGCTCAACAGCTTCAACAGAGAAATATTCACGAAATATCAACAGAAAGCTCTTCT 409     |                 |
| OY  | 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120      |                 |
| Db  | 410 GCCAGTGAAGTCTCTTCCAAGTGAACCTCTCAAGCAATTCGCACTAACATAAACAAGACCT 469     |                 |
| OY  | 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140      |                 |
| Db  | 470 GCTTTAAGCTCGGAGGTGGCTTATCATTTAGCTTACAGCTGGTAGACACGTTAAAAAAG 529       |                 |
| OY  | 141 ProPheIleProIleHisIleMetGluAlaHisIleAlaLeuThrIleArgLeuThrAsnLys 160   |                 |
| Db  | 530 CCATTCATTCCTCATTCATATGAGAGGCTCATGCACTTACTATTGAGTTGACCAATAAA 589       |                 |
| OY  | 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180      |                 |
| Db  | 590 GTAGAAATTCCTTTTAACTTCTTTTGATTTCTGGAGGTCACGCTGTGGGATTAAGTT 649         |                 |
| OY  | 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200      |                 |
| Db  | 650 CAAAGAGTTTCAGATTTTCTGCTCTTGGAAAGCTTTTGGACATACCAACAGGTGACATG 709       |                 |
| OY  | 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisIleProGluCysSerThrMetSer 220   |                 |

|          |                               |   |      |
|----------|-------------------------------|---|------|
| Db       | 710                           | CTTGAACAAGTGGCAAGAAGATTCTTTAATAAAAAATCCAGAGTGCTCAACATGAGT               | 769  |
| Qy       | 221                           | GIYGIYLYSALAIIEGLIHISLEUALALYSGINGLYASNAARGPHEHISPEAPPIELYS             | 240  |
| Db       | 770                           | GGTGGGAAGGCATATGAACATTATGGCCAAACAAAGGAATAATGATTCATTATTTGACATCAA         | 829  |
| Qy       | 241                           | PROPELSEWHISHSIALALYSASNCSYASPPHESERPHETHGILEUGHNHSVALTHR               | 266  |
| Db       | 830                           | CTCCCTTCATCAACAGCTMAAAATATGATGATTTTCTTTACTGAGACTTCAACACGTTACT           | 889  |
| Qy       | 261                           | ASPLYSILEIEMELYSYSGIULYSGIUGIYLIIEGLULDSGLGINILEUSER                    | 280  |
| Db       | 890                           | GATTAATAATATATGAAAAAGCAAAAAGCAAGATTTGAGAAAGGGCAAAATCTGTCT               | 949  |
| Qy       | 281                           | SERIALAIAASPIIEALAIATHRVALGINHSITHMELALCYSHISLEUVALYSARG                | 300  |
| Db       | 950                           | TCAGCAGCAGCATTTGCTGCCACAGTACAGCACACATGCGCATGTCTTGTGAAAAGA               | 1000 |
| Qy       | 301                           | THRHISARGALAIIELEUPHECYSYSGINARASPLEULENPROGLINSENANALAVAI              | 320  |
| Db       | 1010                          | ACACATCCGGCTATTCTGTTTGTGAAGCAGAGACTTGTACTTCAAAATAATGACGTA               | 106  |
| Qy       | 321                           | LEUVALAIASERGIGLYVALAIASERASNPHEYRIIEARGARGALALEUGNIILEU                | 340  |
| Db       | 1070                          | CTGGTTGCATCTGGTGGTGTGCCAAGTAACTTCTATATCCGAGAGCTCTGGAAATTTTA             | 112  |
| Qy       | 341                           | THRAENAIATHRGINCYSSTRLEUENUCYSPROPROAARGIENUCYSTHRAESPANGLY             | 366  |
| Db       | 1130                          | ACAAACGCAACACAGTGACCTTTGTTGTCTCTCCACAGACTATGACATGATAATATGAC             | 118  |
| Qy       | 361                           | IIEMLIEALATTPASNGIYIIEGLUARGLEUARGALIGLYLEUGIYIIELEUHSASP               | 380  |
| Db       | 1190                          | ATTATGATTTGCATGGAATGTGATTTGAAACACTACGCTGGCTTGGGCATTTTATACATGAC          | 124  |
| Qy       | 381                           | IIEGIUGIYIIEARGTGTGLUPROLYSCYSPROLEGLYVALASPIIESEITLYSGIUAL             | 400  |
| Db       | 1250                          | ATAGAAAGGCATCCGCTATGAAACCAAAATGTCCTCTTGAGGTACATATCAAAAGANGTT            | 130  |
| Qy       | 401                           | GIYGIUALASERILEYSAIVPROGLINLEUYSMETGLUIIE                               | 414  |
| Db       | 1310                          | GGAGAAAGCTTCATTAAGTATCACACATTTAAAAATGAGATA                              | 1351 |
| RESULT 2 |                               |   |      |
| ID       | ABST76635                     | standard; DNA; 2197 BP.   |      |
| XX       | ABST76635;                    |   |      |
| XX       |                               |   |      |
| DT       | 11-DEC-2002                   | (first entry)   |      |
| DE       |                               | DNA encoding novel human metalloprotease MPl.                           |      |
| KW       |                               | Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;    |      |
| KW       |                               | motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;    |      |
| KW       |                               | reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;   |      |
| KW       |                               | genital wart; metabolic disorder; premature puberty; Kallman syndrome;  |      |
| KW       |                               | Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;     |      |
| KW       |                               | Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;   |      |
| KW       |                               | liver disease; renal disease; immune disorder; rheumatoid arthritis;    |      |
| KW       |                               | acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia; |      |
| KW       |                               | emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;   |      |
| XX       |                               | neurological disorder; gene; ds.  |      |
| OS       |                               | Homo sapiens.   |      |
| XX       |                               |   |      |
| PN       | WO200272751-A2.               |   |      |
| PD       | 19-SEP-2002.                  |   |      |
| XX       |                               |   |      |
| PF       | 05-FEB-2002; 2002WO-US003351. |   |      |
| XX       |                               |   |      |
| PR       | 05-FEB-2001; 2001US-0266518P. |   |      |

PR 10-APR-2001; 2001US-0282814P.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S;

DR WPI; 2002-723329/78.

DR P-PSDB; ABG6478.

XX New isolated nucleic acid encoding MP-1 protein, useful for preventing, treating, or ameliorating diseases associated with aberrant metalloproteinase activity, e.g. immune, metabolic, inflammatory and neurological disorders.

PS Claim 1; Fig 1A-C; 473pp; English.

XX The invention describes an isolated nucleic acid molecule (I) encoding a metalloproteinase (MP-1). (I) is useful for preventing, treating, or ameliorating a medical condition, particularly an immune disorder, an aberrant glutamate transport or motor neuron disorder, such as amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like condition. The compositions and methods are also useful for diagnosing, prognosticating, treating, ameliorating and/or treating disorders associated with MP-1 activity, e.g. diabetes, cancer, reproductive disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome, or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease, Parkinson's disease, Huntington's disease or Tourette syndrome), liver and renal diseases and immune disorders (e.g. AIDS, Rheumatoid arthritis or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic fibrosis) and vascular, inflammatory and neurological disorders (e.g. Alzheimer's disease or Parkinson's disease). This sequence represents a metalloproteinase MP1 polynucleotide

XX Sequence 2197 BP; 681 A; 441 C; 439 G; 636 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,14e-212 Length: 2197  
Score: 2125.00 Matches: 414  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0

US-10-649-273-2 (1-414) x ABST76635 (1-2197)

OY 1 MetLeuIleuThrIleuThraAlaGlyValPhePheIleuProSerIleuValIleu 20  
DB 231 ATCTAATCTTGAAGTCTGACGAGAGTCTTTTAAACCACTAAAGAAAGTTAT 290  
OY 21 GluPheIleuArgSerPheAsnPheIleuProGlyThrIleuPheIleuIleuVal 40  
DB 291 GAATTTTAAAGAGTTTAAATTTTCATCTCGAACAATTTCTTCAATAAATAGTATG 350  
OY 41 GlyIleuIleuThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60  
DB 351 GGAATTGAAACTGATGCTGATGATACAGACAGCTGCTGGTGGAAGTGAAGAAATGTC 410  
OY 61 LeuGlyGluAlaIleuHisSerGlnThrGluValHisIleuValGlyIleuValPro 80  
DB 411 TTGGAGAAACAAATACATCTCCCAACTGAACTGATTTTAAACAGGTGGAGTTGTTCT 470  
OY 81 ProAlaAlaGlnGlnIleuHisIleuValGlnValIleuValGlnValIleuSer 100  
DB 471 CCAGCACTCAACAGCTTCCACAGAAATTTTCAAGAAATAGTACCAAGAGCTTTCT 530  
OY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleuAlaThrThrIleuProGlyLeu 120  
DB 531 GCCAGTGGAGTCTCTCCAGTGAAGCTCTGACCAATTCACATCAATAAATCTAGACTT 590  
OY 121 AlaLeuSerIleuGlyValGlyLeuSerPheSerIleuGlnIleuValGlyIleuVal 140  
DB 591 GCTTTAAGCTGGAGTGGCTTATATCAATTAGCTTACAGCTGTGAGGACATTAAGAG 650

OY 141 ProPheIleuProIleuHisIleuSerGluAlaHisAlaIleuThrIleuGluThrAsnVal 160  
DB 651 CCATTCAATCCCATTCATCATATGAGAGCTCATGACATCTACTATAGGTACCAATGAA 710  
OY 161 ValGluIleuPheProPheIleuValIleuIleuIleuSerGlyIleuIleuValIleuVal 180  
DB 711 GTAGAAATTTCTTTTATAGTTCTTTTGAATTTCTGAGAGTCACTGCTGTGCACTTACT 770  
OY 181 GlnGlyValSerAspPheIleuIleuGlyIleuSerIleuAspIleuAlaProGlyAspMet 200  
DB 771 CAAGAGTTTCAATTTCTGCTCTTGGAAAGCTTTGGACATGACACAGTGCATG 830  
OY 201 LeuAspIleuValAlaArgArgLeuSerIleuIleuIleuIleuIleuIleuIleuSer 220  
DB 831 CTTCACAAAGTGGCAGAGAGCTTTCTTTAATAAATCCAGAGTGTCCACCATGAGT 890  
OY 221 GlyGlyValAlaIleuGlnIleuValAlaGlnGlnIleuValGlnIleuPheAspIleuVal 240  
DB 891 GGTGGAAAGCCATAGAACATTTGTGCCAAACAGGAAATAGATTTTCATTTTGACATCAA 950  
OY 241 ProProIleuHisIleuAlaValAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
DB 951 CTCTCCCTTCATCATGCTAAATATGATTTTCTTTACTGACCTTCACACCTTACT 1010  
OY 261 AspIleuIleuIleuMetIleuValGlyIleuGlyIleuGlyIleuGlyIleuIleuSer 280  
DB 1011 GATTAATATATATGAAAGAAAGAAAGAAAGAGATATGGAAGGGCAATCTGTCT 1070  
OY 281 SerAlaAlaAspIleuAlaIleuValGlnHisIleuMetAlaCysHisIleuValIleuArg 300  
DB 1071 TCACAGACAGCATTTGCTGCCACAGTACAGACCAATAGGCAATGCTCATTTGTGAAAGA 1130  
OY 301 ThrIleuArgAlaIleuPheCysGlyIleuArgAspIleuValProGlnAsnAlaVal 320  
DB 1131 ACACATCGGCTATTTCTTTTGTAAAGACAGAGACTTTGTAATCTCAATAATATGACATA 1190  
OY 321 LeuValAlaSerGlyIleuValAlaSerAsnPheIleuValIleuValIleuValIleuVal 340  
DB 1191 CTGGTTCATTTGGTGGTGTGCGAAGTACATTTCTTATCCGACAGCTCTGGAATTTTA 1250  
OY 341 ThrAsnAlaThrGlnCysThrIleuIleuCysProProAlaGluCysThrAspAsnGly 360  
DB 1251 ACAACGGAACACAGTGCATTTGTGTGCTCTCCCAAGCTATGACATGATATGAC 1310  
OY 361 IleuIleuAlaIleuAsnGlyIleuGluArgLeuArgAlaGlyLeuGlyIleuHisAsp 380  
DB 1311 ATTATGATTCATGAAATGATGATTAAGAAAGACTACGCTGCTGGCATTTTACATGAC 1370  
OY 381 IleuGluGlyIleuArgGlyIleuProIleuCysProIleuGlyValAlaAspIleuSerGlyVal 400  
DB 1371 ATAGAGGCAATCCGCTATGAAACAAATATGCTCTCTGAGATGACATATCAAAAGAGTT 1430  
OY 401 GlyGluAlaSerIleuValIleuValProGlnIleuIleuIleuIleuIleuIleuIleu 414  
DB 1431 GGAAGAGCTTCATTAAGTACCAATTTAATAAATGAGATA 1472

RESULT 3  
ABST73207  
ID ABST73207 standard; DNA; 2572 BP.  
AC ABST73207;  
DT 01-MAY-2003 (first entry)  
DE Human protein modification + maintenance molecule DNA SEQ ID No 36.  
XX Cytostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;  
XX cerebroprotective; antiparkinsonian; nootropic; antidiabetic; antitumor;  
XX anticancer; hepatotropic; gynaecological; antibacterial; virucide;  
XX protozoacide; antiparasitic; cell proliferative disease; PMOD;  
KM protein modification and maintenance molecule; immunogenic fragment;  
KM cancer; autoimmune; inflammatory disease; neurological disorder;



KW gastrointestinal; developmental; vesicle trafficking disorder; infection;  
KW protein-protein interaction; drug-target interaction;  
KW gene expression profile; human; gene; ds.  
OS Homo sapiens.  
XX WO200300844-A2.  
XX 03-JAN-2003.  
XX 18-JUN-2002; 2002WO-US019360.  
XX 22-JUN-2001; 2001US-0300508P.  
XX 06-JUL-2001; 2001US-0303445P.  
XX 13-JUL-2001; 2001US-0305405P.  
XX 09-AUG-2001; 2001US-0311442P.  
XX 24-AUG-2001; 2001US-0314821P.  
XX 29-AUG-2001; 2001US-0315922P.  
XX 03-MAY-2002; 2002US-0378205P.  
XX (INCY-) INCYTE GENOMICS INC.  
XX Gandhi AR, Kable AE, Swarnakar A, Hafalia AJA, Tran B, Duggan BM,  
PI Warren BA, Ison CH, Honchell CD, Nguyen DB, Lu DAM, Lee EA, Yue H;  
PI Forsythe JD, Barroso I, Ramkumar J, Griffin JA, Li JX, Yang J;  
PI Thangavelu K, Gietzen KU, Ding L, Baugh MR, Borowsky ML, Yao MG,  
PI Wala NK, Mason PM, Gurnajan R, Lee S, Becha SD, Lee SY, Tran UK;  
PI Elliott VS, Luo W, Sprague KW, Tang YT, Lu Y, Zebajadian Y;  
XX WPI; 2003-184039/18.  
XX P-PSDB; ABJ26654.  
XX New isolated human PMOD polypeptide and polynucleotide, useful for  
PT diagnosing, treating and preventing diseases or conditions associated  
PT with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and  
PT infections.  
XX Claim 91; Page 211; 225pp; English.  
XX The invention relates to an isolated polypeptide comprising: any of 28  
CC sequences of 48-1256 amino acids; a natural amino acid sequence at least  
CC 90% identical to the 28 amino acid sequences, 94% identical to a sequence  
CC of 703 or 267 amino acids, 96% identical to a sequence of 414 amino  
CC acids, or 97% identical to a sequence of 242 amino acids, all given in  
CC the specification; or a biologically active or immunogenic fragment of  
CC the isolated polypeptide. The polypeptides and polynucleotides are useful  
CC in diagnosing, treating and preventing diseases or conditions associated  
CC with the decreased expression of protein modification and maintenance  
CC molecules (PMOD), such as cell proliferative diseases (e.g. cancer,  
CC atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,  
CC allergies), neurological disorders (e.g. stroke, Parkinson's disease,  
CC epilepsy), gastrointestinal disorders (e.g. ulcer, cirrhosis), reproductive (e.g.  
CC endometriosis), developmental, vesicle trafficking disorders, and  
CC infections (e.g. bacterial, viral, parasitic, protozoal). These are also  
CC useful in assessing the effects of exogenous compounds on the expression  
CC of nucleic acid and amino acid sequences of PMOD. The PMOD or its  
CC fragments are useful in screening compounds for effectiveness as agonist  
CC or antagonist of the polypeptides, or in altering the expression of the  
CC target polynucleotide and compounds that specifically bind to or modulate  
CC the activity of the polypeptide. The microarray is useful in monitoring  
CC or measuring protein-protein interactions, drug-target interactions, and  
CC gene expression profiles. This polynucleotide sequence represents the DNA  
CC encoding a human PMOD protein of the invention  
XX  
XX Sequence 2572 BP; 780 A; 489 C; 509 G; 794 T; 0 U; 0 Other;  
XX  
XX Alignment Scores:  
XX Pred. No.: 2.7e-212 Length: 2572  
XX Score: 2125.00 Matches: 414  
XX Percent Similarity: 100.00% Conservative: 0  
XX Best Local Similarity: 100.00% Mismatches: 0  
XX Query Match: 100.00% Indels: 0  
XX DB: 8 Gaps: 0

US-10-649-273-2 (1-414) x ABJ23207 (1-2572)  
QY 1 MetLeuIleuThrIleuThrAlaGlyValPhePheIysProSerIysArgIleValTyr 20  
Db 144 ATGCTAATCTTGACTTAAGACTGACAGAGTTTATTTTAAACCACTAAAGAAAGTTTAT 203  
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValIleu 40  
Db 204 GAATTTTAAAGAGTTTAAATTTTCACTCCGGAACACTATTCTTCAATAAATAGTATG 263  
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValIleAspGluThrGlyAsnVal 60  
Db 264 GGAATTAACCTAGTTGTGATGATACAGACAGCTCTGTGTGATGAACCTGGAATGTG 323  
QY 61 LeuGlyIleuAlaIleHisSerGlnThrGluValHisIleuIysThrGlyIleValPro 80  
Db 324 TTGGAGACAGCAATACATTCCTCAAACTGAAGTTATTTAAACAGGAGGAGTTTCT 383  
QY 81 ProAlaIleGlnIleuHisArgGluAsnIleGlnArgIleValGlnIleuAlaLeuSer 100  
Db 384 CCAAGCACTCAACAGCTTCCACAGAAATTTTCAAGAAATGACAAAGAGCTTTCT 443  
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleIysProGlyLeu 120  
Db 444 GCCAGTGAAGTCTCCCAAGTGAAGTCTCCAGCAATTCACATCAATCAACAGAGACTT 503  
QY 121 AlaLeuSerLeuGlyValIleGlyLeuSerPheSerLeuGlnIleuValIleGlnIleuIys 140  
Db 504 GCTTTAAGCCTGGAGTGGCTTATTCATTTAGCTTACAGCTGTGAGCACTTTAAAG 563  
QY 141 ProPheIleProIleHisIleMetGluAlaHisIleAlaLeuThrIleArgLeuThrAsnIys 160  
Db 564 CCAATTCATTCATTCATATGAGGCTCATCTTACTATTAAGTTGACCAATAA 623  
QY 161 ValGluPheProPheLeuValIleuLeuIleSerGlyGlnHisCysLeuLeuAlaLeuVal 180  
Db 624 GTAAATTTCTTTTATTTAGTTCTTTTATTTCTGAGGTCACTGTGTGCAATTAAT 683  
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyIysSerLeuAspIleAlaProGlyAspMet 200  
Db 684 CAAGAGCTTTCAGATTTTCTGCTTCTTGAAGCTTTGGAACATGACCAAGTCACTG 743  
QY 201 LeuAspIysValAlaArgArgLeuSerLeuIleIysHisProGluCysSerThrMetSer 220  
Db 744 CTGACAGGTGGACAGAGACCTTTCTTAAATAACATCCAGAGTCTCCACATGAGT 803  
QY 221 GlyIleIysAlaIleGlnHisIleuAlaIysGlnIysAsnArgPheHisPheAspIleIys 240  
Db 804 GGTGGAAAGCATAGAAACAATTTGGCCAAACAGAAATATGATTTCAATTTGAACATCAA 863  
QY 241 ProProLeuHisIleAlaIysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
Db 864 CTTCTCCATGATCAATGCTAAATAATTTGATTTCTTTTACCTGACCTTCAACAGCTTACT 923  
QY 261 AspIysIleIleMetIysGlnIysGluGlnIleGlnIleGlnIleGlnIleIleuSer 280  
Db 924 GATTAATAATTAATGAAGAAAGAAAGAAAGAGATATGAAAGGGCAATCTGTCT 983  
QY 281 SerAlaIleAspIleAlaIleAlaThrValGlnHisIleThrMetAlaCysHisIleValIysArg 300  
Db 984 TCAGACACAGACATTCCTCCACAGTACAGACACATGCAATGATGATCTTGTGAAAGA 1043  
QY 301 ThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320  
Db 1044 ACAATCGGGCTATTTCTTTTGTATGACAGAGACTTTGATCTCAAAATTAATGACATA 1103  
QY 321 LeuValAlaSerGlyValAlaIleSerAsnPheThrIleArgAlaIleGlnIleu 340  
Db 1104 CTGTTTCATCTGTGTGTGTGTGCAAGTAATCTTCAATTCACAGAGCTCTGAAATTTTA 1163  
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaGluCysThrAspAsnGly 360

|          |      |   |  |
|----------|------|---|--|
| Db       | 1164 | ACAAAGCAACAGACGCTTGTGTGTCTCCAGACTATGACATGATTAATGCG                            | 122  |
| Qy       | 361  | 11etwct11eal1atTPaeng1y11egiua1g1eua1g1a1g1y11e1u1h1sa1p                      | 380  |
| Db       | 1224 | ATTATGATTCATCGAATGTGATTGAAAGACTACGTCGGCTTGGCATTTTACATGAC                      | 128  |
| Qy       | 381  | 11egiug1y11a1a1r1g1y11u1p1r1o1y1s1c1p1r1e1u1g1v1a1a1s1p11e1s1e1r1y1s1g1u1a1   | 400  |
| Db       | 1284 | ATGAGAGGATCCGCTATATACCAAAATCTCTTGGAGTACATATCCAAAGAGTT                         | 134  |
| Qy       | 401  | GIYGIUA1a1s1e1r11e1y1s1v1a1P1r1o1g11e1u11e1y1s1m1e1g1u11e                     | 414  |
| Db       | 1344 | CGAGAGCTTCATTAAGTACACACATTAATAAGATGAGATA                                      | 1385   |
| RESULT 4 |      |   |  |
| AAD46856 | ID   | AAD46856 standard; cDNA; 1820 BP.   |  |
| AC       | AC   | AAD46856;   |  |
| XX       | XX   |   |  |
| DT       | DT   | 27-JAN-2003 (first entry)   |  |
| DE       | DE   | Human glycoprotease 28472 cDNA.   |  |
| XX       | XX   |   |  |
| KW       | KW   | Human, adenosine deaminase; seven transmembrane domain receptor; cancer;      |  |
| KW       | KW   | 77M; glycoprotease; immune disorder; IGA deficiency; allergy; arrhythmia;     |  |
| KW       | KW   | rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular disorder;     |  |
| KW       | KW   | hypercholesterolemia; ischemic heart disease; obesity; myocardial infarction; |  |
| KW       | KW   | endothelial cell disorder; Grave's disease; psoriasis; brain disorder;        |  |
| KW       | KW   | Parkinson's disease; Alzheimer's disease; haematopoietic disorder;            |  |
| KW       | KW   | peripheral oedema; metabolic disorder; liver disorder; platelet disorder;     |  |
| KW       | KW   | chromosome mapping; tissue typing; gene therapy; neuroprotective;             |  |
| KW       | KW   | cytostatic; anorectic; cardiac; haemostatic; gene; ss.                        |  |
| OS       | OS   | Homo sapiens.   |  |
| XX       | XX   |   |  |
| EH       | EH   | Key   | Location/Qualifiers  |
| FT       | FT   | CDS   | 146..1390  |
| FT       | FT   |   | /tag= a  |
| FT       | FT   |   | /product= "Human 28472 protein"  |
| FT       | FT   |   | /note= "This region is specifically claimed as SEQ ID NO: 6 in claim 1 of the specification" |
| XX       | XX   |   |  |
| PN       | PN   | WO200274960-A2.   |  |
| XX       | XX   |   |  |
| PD       | PD   | 26-SEP-2002.  |  |
| XX       | XX   |   |  |
| PP       | PP   | 08-NOV-2001; 2001WO-US051427.   |  |
| XX       | XX   |   |  |
| PR       | PR   | 08-NOV-2000; 2000US-0246768P.   |  |
| PR       | PR   | 08-NOV-2000; 2000US-0246772P.   |  |
| PR       | PR   | 15-NOV-2000; 2000US-0249185P.   |  |
| XX       | XX   |   |  |
| PA       | PA   | (MILL-) MILLENNIUM PHARM INC.   |  |
| XX       | XX   |   |  |
| PI       | PI   | Leibyl KR, Kapeller-libermann R, Glucksmann M;                                |  |
| XX       | XX   |   |  |
| DR       | DR   | WPI: 2002-759898/82.  |  |
| DR       | DR   | P-PSDB; AAE29234.   |  |
| XX       | XX   |   |  |
| PT       | PT   | New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,         |  |
| PT       | PT   | useful for diagnosing and treating cancer, immune, cardiovascular,            |  |
| PT       | PT   | hematopoietic, brain, pain, metabolic, liver or platelet disorders, and       |  |
| PT       | PT   | in pharmacogenomics.  |  |
| XX       | XX   |   |  |
| PS       | PS   | Claim 1, Fig 8, 178pp; English.   |  |
| XX       | XX   |   |  |
| CC       | CC   | The present invention relates to novel 38650, 28472, 5495, 65507, 81588       |  |
| CC       | CC   | or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-      |  |
| CC       | CC   | protease or seven transmembrane domain (77M) receptor family members.         |  |
| CC       | CC   | Sequences of the invention are useful in diagnosing and treating cancer       |  |
| CC       | CC   | or aberrant cellular proliferation and/or differentiation (e.g. colon or      |  |
| CC       | CC   | lung cancer), immune disorders (e.g. selective IGA deficiency, rheumatoid     |  |

|    |   |   |               |      |
|----|---|---|---------------|------|
| CC | arthritis, diabetes, allergy), heart or cardiovascular disorders. (e.g., hypertension, atherosclerosis, arrhythmias, ischaemic heart disease, myocardial infarction, thrombus) including endothelial cell disorders (e.g. psoriasis, Grave's disease), haematopoietic disorders, brain disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease), pain and metabolic disorders (e.g. obesity), liver disorders or platelet disorders. They are also useful in screening assays, predictive medicine (e.g. diagnostic assays, prognostic assays, monitoring clinical trials and pharmacogenetics) and prophylactic and therapeutic methods. The nucleic acids may also be used in chromosome mapping, tissue typing and forensic biology and as surrogate markers. Sequences of the invention are also used in gene therapy. The present sequence is human glycoprotease 28472 CDNA |   |               |      |
| CC | XX  | Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other; |               |      |
| CC | US-10-649-273-2 (1-414) x AAD46856 (1-1820)   |   |               |      |
| CC | Alignment Scores:   |   |               |      |
| CC | Pred. No.:  | 1.24e-208   | Length:       | 1820 |
| CC | Score:  | 2088.00   | Matches:      | 407  |
| CC | Percent Similarity:   | 99.03%  | Conservative: | 3    |
| CC | Best Local Similarity:  | 98.31%  | Mismatches:   | 4    |
| CC | Query Match:  | 98.26%  | Indels:       | 0    |
| CC | DB:   | 6   | Gaps:         | 0    |
| CC | 1 MetLeuIleLeuThrLysThraAlaGlyValAPhePheLysProSerLysArgLysValTyr 20   |   |               |      |
| CC | 146 ATGCTAACTTGCACCTACAGAGCTCAGAGAGTTTAAAAACAATAAGAAAGTTAT 205  |   |               |      |
| CC | 21 GluPheLeuArgSerPheAsnPheHisProGlyLysPheLeuHisLysIleValLeu 40   |   |               |      |
| CC | 206 GAATTTTAAAGAGTTTATTTTCAATCTCGAACACTATTTCTCATATAAATAGATTG 255  |   |               |      |
| CC | 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60   |   |               |      |
| CC | 266 GGAATTGAACACTAGTTGTGATATACAGACGCTCTGTGTGATGATAACTGAAATGTG 325   |   |               |      |
| CC | 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80   |   |               |      |
| CC | 326 TTGGAGAGCAATACATTCCTCCAACTGAAAGTTCAATTAATAACAGGTGGATTGTCCT 385  |   |               |      |
| CC | 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100   |   |               |      |
| CC | 386 CCAGAGCTCAACAGCTTACAGAGAAATATTCACGAATACTACAGAGCTCTTCT 445   |   |               |      |
| CC | 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  |   |               |      |
| CC | 446 GCCATGAGAGTCTCTCCAAAGTACCTCAGCAATTCGAACATACATATAAACAGAGCTT 505  |   |               |      |
| CC | 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140   |   |               |      |
| CC | 506 GCTTTAAGCTGGAGTGGCTTATCATTTAGCTTACACTGTGTAGGACAGTTAAAAAG 565  |   |               |      |
| CC | 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  |   |               |      |
| CC | 566 CCATTTCATTCCTCATATCATATGAGGCTCAAGCACTTACATTAATAGTGTACCAATAAA 625  |   |               |      |
| CC | 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  |   |               |      |
| CC | 626 GTAGAAATTCCTTTTATGATCTCTTTGATTTTCGAGAGTCACACTGTCTGTGGCATTAAGT 685   |   |               |      |
| CC | 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  |   |               |      |
| CC | 686 CAAAGAGTTTCAGATTTTCTGCTTCTTGAAAGTCTTTGGACATAGCACAGAGTGACATG 745   |   |               |      |
| CC | 201 LeuAspLysValAlaAlaArgArgLeuSerLeuIleLysHisProGluLysSerThrMetSer 220   |   |               |      |
| CC | 745 CTGTGACACCTCCCAACAACAATCTTTTAAATAAATCTCCAGAGTCTCCACCATAGGT 805  |   |               |      |
| CC | 221 GlyGlyLysValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240   |   |               |      |
| CC | 806 GGTGGAAAGCCATATGAACTTTGGCCAAACAAGAAATATGATTTCAATTTTGACATCAAA 865  |   |               |      |

QY 241 ProProLeuHsihsalalysaenCyaspheSerPheThrglyLeuGlnhisvalThr 260  
 DB 866 CCTCCCTTCATCATCTAAATAATGATTTCTTTTACTGCACTTCAACAGTACT 925  
 QY 261 AspLysIleIleleuetylsylsGluLysGluGluGlyIleGluLysGlyGlnIleLeuSer 280  
 DB 926 GATTAATAATATGATAAAGAGGAAAGAGGAAAGATTAAGAGGAGGAAATCTGTCT 985  
 QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300  
 DB 986 TCAGCAGCAGACATCTGCTGCCACAGTACAGACACATGCAATGCAATGCTATCTTGGAAAGA 1045  
 QY 301 ThrHisArgAlaAlaIleLeuPheCysLysGlnArgAspLeuPheProGlnAsnAlaVal 320  
 DB 1046 ACACATCGGGCTATCTGTTTGAAGCAGAGACTGTTTACTCAAAATATATGCACTA 1105  
 QY 321 LeuValAlaSerGlyGlyValAlaAlaSerAspPheTyrIleArgArgAlaLeuGluIleLeu 340  
 DB 1106 CTGGTTCATCTGGTGGTGTGCGAAGTAATCTCTATATCCGACAGCTCTGGAATTTTA 1165  
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360  
 DB 1166 ACAAGCGAACACAGTGCATTTGTTGTGTCTCTCCAGACTATGCACTGATATATGAC 1225  
 QY 361 IleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
 DB 1226 ATTATGATTTGCATGCAATGCTATTTGAAGACTACGCTGCTGGCATTTTATCAGTAC 1285  
 QY 381 IleGluGlyIleArgTyrGluProLysCysProLeuGlyValAlaAspIleSerLysGluVal 400  
 DB 1286 ATAGAAGGATCGCTGATGAAACCAAAATGTCCTTGGAGTACATATCAAAAGAGTT 1345  
 QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414  
 DB 1346 GGAGAGCTTCATTAATAAGTACCAATTAATAAGAGATA 1387

RESULT 5  
 ID ACA60887 standard; cDNA; 1820 BP.  
 AC ACA60887;

08-JUL-2003 (first entry)

Human cDNA 28472 encoding a glycoprotease.  
 Human; ss; gene; cancer; aberrant cellular proliferation;  
 differentiation; immune disorders; heart disorder; brain disorder;  
 cardiovascular disorder; endothelial cell disorder; pain disorder;  
 haematopoietic disorder; blood vessel disorder; metabolic disorder;  
 liver disorder; platelet disorder; glycoprotease.

Homo sapiens.

Key Location/Qualifiers  
 CDS 146..1390  
 /tag= a  
 /product= "Glycoprotease"  
 /note= "This CDS is specifically claimed in claim 1"

US2003009017-A1.

09-JAN-2003.

08-NOV-2001; 2001US-00012140.

08-NOV-2000; 2000US-0246768P.  
 08-NOV-2000; 2000US-0246772P.  
 15-NOV-2000; 2000US-0249185P.

(LEIB/) LEIBY K R.  
 (KAP/) KAPPELLER-LIBERMANN R.

PA (GLUC/) GLUCKSMANN M A.  
 PI Leiby KR, Kapeller-Libermann R, Glucksmann MA;  
 XX WPI: 2003-42888/40.  
 DR P-PSDB: AB009569.  
 DR New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid  
 PT molecules, useful for diagnosing, treating cancer, pain, or immune,  
 PT heart, endothelial cell, hematopoietic, blood vessel, brain, metabolic  
 PT and liver disorders.

Claim 2, Fig 8; 90pp; English.

The invention relates to an isolated 38650 (encoding adenosine  
 CC deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7  
 CC transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or  
 CC a sequence which is at least 60% identical to the six nucleic acids or  
 CC their open reading frames, fragments of at least 15 nucleotides,  
 CC naturally occurring variants, or a DNA insert of the plasmid deposited in  
 CC with the American Type Culture Collection as Accession No. not defined in  
 CC the specification, which encodes the amino acid sequence). Also included  
 CC are a host cell containing the nucleic acids (used to produce the  
 CC proteins), the encoded proteins, an antibody that selectively binds to  
 CC the polypeptide, and identifying a compound that binds to/modulates the  
 CC activity of the polypeptide. The nucleic acid molecules, polypeptides and  
 CC methods are useful for diagnosing, treating cancer, aberrant cellular  
 CC proliferation and/or differentiation, immune disorders, heart disorders,  
 CC cardiovascular disorders including endothelial cell disorders,  
 CC haematopoietic disorders, blood vessel disorders, brain disorders, pain  
 CC and metabolic disorders, liver disorders and platelet disorders (many  
 CC examples of these disorders are given in the specification). The present  
 CC sequence is the Human cDNA 28472 encoding a glycoprotease

SQ Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.: 1,24e-208 Length: 1820  
 Score: 2088.00 Matches: 407  
 Percent Similarity: 99.03% Conservative: 3  
 Best Local Similarity: 98.31% Mismatches: 4  
 Query Match: 98.26% Indels: 0  
 DB: Gaps: 0

US-10-649-273-2 (1-414) x ACA60887 (1-1820)

QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20  
 DB 146 ATGCTAATCTTGACTTAAGACTGCGAGAGTTTATTAACCATCAAAAGAAAGTTTAT 205  
 QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40  
 DB 206 GAATTTTAAGAACTTTTAATTTTCATCTGGAACACATATTTCTCATTAATAATGATATG 265  
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaAlaValLysAspGluThrGlyAsnVal 60  
 DB 266 GGAATTAAGAACTAATTTGATGATACACAGCTGCTGCGATGAACTGAAATG 325  
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80  
 DB 326 TTGGAGAGAGCAATTAATTCCTCAAACTGAAGTTCATTTAAACAGGCGGATGTTCT 385  
 QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
 DB 386 CCAAGCACTCAACAGCTTCACAGAGAAATTTTACAGCAATATGACAGAGCTCTTCT 445  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
 DB 446 GCCAGTGAAGTCTTCCAGAGACCTCTCAGCAATTCACATTAACACAGAACTT 505  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140  
 DB 506 GCTTTAAGCTCGGAGAGGCTTATTCATTTAGCTTAACGCTGATAGACAGTAAAG 565

QY 141 ProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
 DB 566 CCATTATCCCATTCATCATATGAGAGCTCATGCACCTTACTATTAGGTGACCAATATA 625  
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisGlyLeuLeuAlaLeuVal 180  
 DB 626 GTAGAAATTCCTTTTATGATCTTTGATTTCTGGAGGTCACTGCTGTTGGCACTTATGTT 685  
 QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
 DB 686 CAAAGGATTCAGATTTTCTGCTTCTTGAAAGCTTTGGACATAGCAACAGGTGACATG 745  
 QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220  
 DB 746 CTTGACAGGTGGCAAGAAACATTTCTTTATATAAACATCCAGAGTCTCCACCATAGT 805  
 QY 221 GlyGlyLysAlaIleGlyHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
 DB 806 GGTGGGAAAGCCATAGAACATTTGGCCAAACAAAGAAATAGATTTCAATTTTGACATCAA 865  
 QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
 DB 866 CCTCCCTTGATCATGCTATAAAATGATTTTCTTTTACTGACTGCAACACGTTACT 925  
 QY 261 AspLysIleIleMetLysLysGlyLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 280  
 DB 926 GATTAATAATATATAAAACAGAAACAGAAAGAGTATTGAGAAAGGAGGCAATCTGCT 985  
 QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLysValLysArg 300  
 DB 986 TACGACAGCAGACATTTGCTGCACAGTACAGACACAAATGGCATGCTCTTGATAAGA 1045  
 QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320  
 DB 1046 ACACATCGGGCTATTCGTTTGTGTAACAGAGACATTTGTAACCTCAAAATAAGCAGTA 1105  
 QY 321 LeuValAlaSerGlyValAlaAlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeu 340  
 DB 1106 CTGGTGCATCTGTGTGTGCAAGTAACTTCTTAATCCGACAGCTCTGGAATTTTA 1165  
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgArgLeuCysThrAspAsnGly 360  
 DB 1166 ACAAGCGCAACAGTCACTTTGTTGTGCTCTCTCCAGACATATGACATGATATGCG 1225  
 QY 361 IleMetIleAlaThrAsnGlyIleGlyArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
 DB 1226 ATTATGATTCATGAGATGGAATGGAATGGAAGACTAGCTGCTGGCATTTTACATGAC 1285  
 QY 381 IleGlyGlyIleArgGlyGlyProLysCysProLeuGlyValAlaAspIleSerLysGlyVal 400  
 DB 1286 ATGAAAGGCTCCGCTATGAAACCAAAATGCTCTTGGAGTAGACATATCAAAAGAGTT 1345  
 QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414  
 DB 1346 GGAAGAGCTTCATTAAGTACCAATTAATAATGAGATA 1387  
 RESULT 6  
 ABSS57020  
 ID ABSS57020 standard; cDNA; 1821 BP.  
 AC ABSS57020;  
 DT 30-JAN-2003 (first entry)  
 XX cDNA encoding novel human glycoprotease 28472.  
 DE  
 KW Cancer; aberrant cell proliferation; aberrant cell differentiation;  
 KW breast cancer; ovarian cancer; prostate cancer; colon cancer;  
 KW lung cancer; immune disorder; heart disorder; cardiovascular disorder;  
 KW endocervical disorder; hematopoietic disorder; blood vessel disorder;  
 KW brain disorder; pain; metabolic disorder; liver disorder; diabetes;  
 KW platelet disorder; carcinoma; sarcoma; leukaemia; Hodgkin's disease;

KW autoimmune disorder; hypertension; atherosclerosis; heart failure;  
 KW myocardial infarction; ischaemic heart disease; Crohn's disease;  
 KW Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;  
 KW cerebral ischaemia; peripheral neuropathy; Alzheimer's disease;  
 KW Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; gene; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 147..1391  
 FT /tag=a  
 FT /product="Glycoprotease 28472"  
 FT /note="Specifically claimed in claim 1"  
 FT  
 FN WO200272233-A2.  
 XX  
 XX 03-OCT-2002.  
 XX  
 XX 08-NOV-2001; 2001WO-US046724.  
 XX  
 XX 08-NOV-2000; 2000US-0246768P.  
 XX 08-NOV-2000; 2000US-0246772P.  
 XX 15-NOV-2000; 2000US-0249185P.  
 XX  
 XX (MILL-) MILLENNIUM PHARM INC.  
 PA  
 PI Leiby KR, Kapeller-Libermann R, Glucksmann M;  
 XX WPI; 2003-029938/02.  
 DR P-PSDB; ABG71162.  
 XX  
 XX New adenosine deaminase, glycoprotease and seven transmembrane domain  
 PT nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,  
 PT 81588 and 14354, useful for treating e.g. leukemias, Hodgkin's disease or  
 PT hypertension.  
 PS  
 PS Claim 2; Fig 8A-B; 178bp; English.  
 XX  
 XX The invention describes 38650, 28472, 5495, 65507, 81588 and  
 CC 14354 nucleic acid molecules (I) and their encoded polypeptides (II). The  
 CC 38650 nucleic acid molecule comprises a sequence encoding adenosine  
 CC deaminase. The 28472 nucleic acid molecule comprises a sequence encoding  
 CC a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise  
 CC sequences that encode a human seven transmembrane domain (7TM). The  
 CC 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide  
 CC sequences are useful for diagnosing, preventing or treating a subject  
 CC with or at risk of developing a disorder, e.g. cancer or aberrant  
 CC cellular proliferation and/or differentiation (e.g. breast, ovarian,  
 CC prostate, colon or lung cancer), immune disorders, heart disorders,  
 CC cardiovascular disorders, endocervical disorders, hematopoietic disorders,  
 CC blood vessel disorders, brain disorders, pain and metabolic disorders,  
 CC liver disorders or platelet disorders. These disorders include carcinoma,  
 CC sarcoma, leukaemia, Hodgkin's disease, autoimmune disorders,  
 CC hypertension, atherosclerosis, heart failure, myocardial infarction,  
 CC ischaemic heart disease, Crohn's disease, Grave's disease, Kawasaki  
 CC syndrome, Raynaud's disease, aneurysm, cerebral ischaemia, peripheral  
 CC neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,  
 CC cachexia or diabetes. This sequence encodes the novel human glycoprotease  
 CC 28472  
 XX  
 XX Sequence 1821 BP; 543 A; 365 C; 394 G; 518 T; 0 U; 1 Other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 1.24e-208 Length: 1821  
 Score: 2088.00 Matches: 407  
 Percent Similarity: 99.03% Conservative: 3  
 Best Local Similarity: 98.31% Mismatches: 4  
 Query Match: 98.26% Indels: 0  
 DE: 10 Ctps: 6  
 US-10-649-273-2 (1-414) x ABSS57020 (1-1821)  
 QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20

```

Db      147 AGCTAAATCTTGAAGCTAGACGAGAGGTTTTTTTAAACATCAAAAGAAAGTTAT 206
Qy      21 GIupheLeuArgSerPheasnPhehi:SPROGLYThrLeuPheLeuHisIleValLeu 40
Db      207 GAATTTTAAAGAGTTTAAATTTTCACTCTGGAACCTATTTCTTCAATAAATAGATATG 266
Qy      41 GIYIleGIuThiSerCysaspPheThraIaIaIaValaValaAspGIuThrGIYasnVal 60
Db      267 GGAATTTGAAGACTAGTGTATGATATACAGAGCTGCTGTGTGTATGTAAGAACTGGAATGTC 326
Qy      61 LeuGIYGIuAlaIleHisSerGIuThrGIuValHisLeuLysThrGIYGIYIleValPro 80
Db      327 TTGGGAGAGCAATACATTTCCCAACCTGAAGTCAATTTAAAAACAGGTGGATGTTCTCT 386
Qy      81 ProAlaIaGIuGIuLeuHisArgGIuAsnIleGIuArgIleValaGIuGIuAlaLeuSer 100
Db      387 CCAAGAGCTCAACAGCTTCAACAGAAATATTCACAGAAATAGTACAGAAAGCTCTTTCT 446
Qy      101 AlaSerGIYValSerProSerAspLeuSerAlaIleAlaThrIleLysProGIYLeu 120
Db      447 GCGAGTGGAGTCTCTCCAGTATGACCTCTCAGCAATTCGCACTACCAATAAACCAGAGCTT 506
Qy      121 AlaLeuSerLeuGIYValGIYLeuSerPheSerLeuGIuLeuValGIYGIuLeuLysLys 140
Db      507 GCTTTAAGCTGGAGAGTGGCTTATCATTTAGCTTACAGCTGGTAGAGACGTTAAAAAG 566
Qy      141 ProPheIleProIleHisIleMetGIuAlaHisIleLeuThrIleArgLeuThrAsnLys 160
Db      567 CCAATTCATTTCCATTCATCATATATGAGAGCTCATGACCTTATCATTTAGTTCACCAATAA 626
Qy      161 ValGIuPheProPheLeuValLeuLeuIleSerGIYGIYHisCysLeuLeuAlaLeuVal 180
Db      627 GTAGAAATTTCTTTTATGTTCTTTGATTTCTGGAGTACTGTCTGTTGGCATTTAGTT 686
Qy      181 GIuGIYValSerAspPheLeuLeuLeuGIYLysSerLeuAspIleAlaProGIYAspMet 200
Db      687 CAAGAGATTTCAAGATTTTCTGCTTTGGAAAGCTTTTGACATAGACACAGAGGATGATG 746
Qy      201 LeuAspLysValAlaIaArgArgLeuSerLeuIleLysHisSPROGLYCysSerThrMetSer 220
Db      747 CTTGACAGAGTGGCAAGAGACTTCTTTAATAAATCATCCAGAGTCTCCACCATAGT 806
Qy      221 GIYGIYValAlaIleGIuHisLeuAlaLysGIuGIYAsnArgPheHisPheAspIleLys 240
Db      807 GGTGGAAAGCCATAGACATTTGGCCAAACAGAAATGATTTTCACTTTTGACATCAAA 866
Qy      241 ProProLeuHisIleHisAlaLysAsnCysAspPheSerPheThrGIYLeuGIuHisValThr 260
Db      867 CCTCCCTTCATCATGCTAAATAATTTGATTTTCTTTTACTGAGACTTCAACACGTTACT 926
Qy      261 AspLysIleIleMetLysLysGIuLysGIuGIYIleGIuLysGIYGIuIleLeuSer 280
Db      927 GATTAATAATATGAAACAGAAACAAAGAGGATTTATGAAAGGCGCAAACTCTGTCT 986
Qy      281 SerAlaIaAspIleAlaIaThrValGIuHisThrMetAlaCysHisIleValLysArg 300
Db      987 TCAAGACGACATCTCTGCAAGTACGACACACAAATGCAATGCAATCTTGTGAAGA 1046
Qy      301 ThrHisArgAlaIleLeuPheCysLysGIuArgAspLeuLeuProGIuAsnAsnAlaVal 320
Db      1047 ACAACATCGGCTATTTCTGTTTGAACAGAGACTTTTACTCTCAAAATATATGACGTA 1106
Qy      321 LeuValAlaIleSerGIYGIYValAlaIaSerAsnPheThrIleArgArgAlaLeuGIuIleLeu 340
Db      1107 CTGGTTCATCTGGTGTGTGCGCAAGTAACTTCTATATCGCAGAGCTCTGGAATTTTA 1166
Qy      341 ThrAsnAlaThrGIuCysThrLeuLeuCysProProArgArgLeuCysThrAspAsnGIY 360
Db      1167 ACAAAAGCAACACAGAGCACTTTGTGTCTCTCCAGACTATGCACTGATTAAGGC 1226
Qy      361 IleMetIleAlaIleThraPasnGIYIleGIuArgLeuArgAlaGIYLeuGIYIleLeuHisAsp 380

```

```

Db      1227 ATTATGATTCGANGCAAGTGGTATTGAAAGACTACGTCTGGCTTGGCATTTTACATGAC 1286
Qy      381 IleGIuGIYIleArgTYrGIuProLysCysProLeuGIYValaAspIleSerLysGIuVal 400
Db      1287 ATGAAAGCATCCGCTATGAAACCAAAATGTCTCTTGGAGTAGACATATCAAAAGAACTT 1346
Qy      401 GIYGIuAlaSerIleLysValProGIuLeuLysMetGIuIle 414
Db      1347 GGAAGAGCTTCATTAAGTACCAATTAATAATGAGATTA 1388

RESULT 7
ADA52832
ID ADA52832 standard; cDNA; 2208 BP.
XX
AC ADA52832;
XX
DE 20-NOV-2003 (first entry)
XX
DE Human coding sequence, SEQ ID 400.
XX
XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
OS Homo sapiens.
XX
PN BP1293569-A2.
XX
PD 19-MAR-2003.
XX
PF 21-MAR-2002; 2002EP-00006586.
XX
PR 14-SEP-2001; 2001JP-00328381.
PR 24-JAN-2002; 2002US-0350435P.
XX
XX (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
XX
DR WPI: 2003-395539/38.
DR P-PSDB: ADA54471.
XX
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 1; SEQ ID NO 400; 205pp; English.
XX
CC The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 2208 BP; 660 A; 454 C; 451 G; 643 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,15e-193 Length: 2208
Score: 1944.00 Matches: 386
Percent Similarity: 93.24% Conservative: 0
Best Local Similarity: 93.24% Mismatches: 4
Query Match: 91.48% Indels: 24
DB: 10 Gaps: 1

US-10-649-273-2 (1-414) x ADA52832 (1-2208)
Qy      1 MetLeuIleLeuThrIleLysThraIaGIYValaPhePheLysProSerLysArgLysValTYr 20
Db      344 ATGCTAAATCTTGACATAAGACTGACAGAGTGTATTTTAAACCATCAAAAGAAAGTTAT 403

```

QY 21 GIUPELEUARGSERPHEASNPHEHISPROGLYTHREUPHEUHSISLYSILEVALEU 40  
DB 404 GAATTTTTTAAAGAGTTTATTTTCAATCCGGAACACTATTTCTCTCAATAAATAGATTTG 463  
QY 41 GIVILEGLUTherSerCySaAspThrAlaAlaValaAspGluThrGlyAsnVal 60  
DB 464 GGAATTGAACTGATGTGATGATACAGCAGCTGCTGTGTGTGATGAACTGGAAATGTG 523  
QY 61 LEUGLYGLUALAILEHISSEGLINTHRGLUVALHISLEUYSERTHRGLYGLILEVALPRO 80  
DB 524 TTGGGAGAGCAATACATTCCTCCAACTGAAGTTCAATTAATAACAGTGGGATTTGTTCT 583  
QY 81 PROALAALAGLNGLEUHSISARGJUVASNILEGNARGILEVALGINGLUVALLEUSER 100  
DB 584 CCAGCAGCTCAACAGCTTCACAGAGAAATATTCAGAAATAGACAGAGCTCTTTCT 643  
QY 101 ALASERGLYVALSERPROSERASPLEUSERAlaIleAlaThrThrIleYSPROGLYLEU 120  
DB 644 GGCAGTGGAGTCTCTCCAACTGACCTCTCAGCAATTGCACTACCAATMAAACGAGACTT 703  
QY 121 ALALEUSERLEUGLYVALGILEUSERPHESEGLINLEUVALGILEUVALYS 140  
DB 704 GCTTTAAGCTGGAGAGTGGCTTATCATTTAGCTTACAGCTGGTAGCAGTTAAAAAG 763  
QY 141 PROPHLEIPEROIIEHISMETGLUALAHISALALEUTHRILEARGLUETHRASNLYS 160  
DB 764 CCATTCATTCCTCATTCATCATATGAGAGCTCATGCACTTACTATAGTTGACCAATAAA 823  
QY 161 VALGIUPHAPROPHLEUVALLEUULISERGLYGLYHISCYLSLEULEUALALEUVAL 180  
DB 824 GTAGAAATTCCTTTTATGTTCTTTGATTTCTGAGAGTCACTGTCTGTGGCATTTAGTT 883  
QY 181 GINGLYVALSERASPHLEULEULEUGLYLYSERLEUSPILAEIAPROGLYASPMET 200  
DB 884 CAAGGAGTTTCAATTTCTGCTTTGGAAGCTTTGACATAGACCAAGGAGCATG 943  
QY 201 LEUASPLYVALAALARARGLEUSERLEULIEYHISPROGLUCYSETHMETSER 220  
DB 944 CTTGACAAAGTGCAGAGACATTCCTTTAATAAACAATCCAGAGTCTCCACCATAGT 1003  
QY 221 GJGLYLYENALILEGLUHSISLEUALYSGINGLYASNARGPHEHISPEASPILEY 240  
DB 1004 GGGGGAGAAAGCCATAGAACATTTGGCCAAACAGAAATGATTTGATTCATCAATA 1063  
QY 241 PROPHLEUHSISALALYASNQYASPHSESERPHEHISGLYLEUHSISVALTHR 260  
DB 1064 CCTCCTTGATCATGTCTAAATAATGTGATTTCTTTTACTGACATTCACACAGTTACT 1123  
QY 261 ASPLYLEILEMETLYSGLULYSGJUGLYGILEGLULYSGJUGLYGILEUSER 280  
DB 1124 GATMAAATAATATGAAAAAGAAAAAGAGAGATTTGAGAAAGGAGCAATCTGTCT 1183  
QY 281 SERIALASPILEALALATHRVALGINHISHERMICALCYSHISLEUVALYARG 300  
DB 1184 TCGGCGACACATTCGTGCGACAGTACAGACACACATTCGATTCATCTTGTAAGAA 1243  
QY 301 THHISARGALALEUPHECYLYSGJNARGASPLEUEUPROGINASNAALAVAI 320  
DB 1244 ACGCATCGGGCTATTCTTTTGTAGCAGAGACTTGTACTCAAAATATATGACGTA 1303  
QY 321 LEUVALALASERGLYGLYVALALASERASNPHETRILEARAGALALEUGLUILEU 340  
DB 1304 CTGGTTCATCTGTGTGTGCGCAAGTAACTTCTGTATCCGCAAGCTCTGGAATTTTA 1363  
QY 341 THASNALATHRGINCYSTRILEUENQYSPROPROKARGLEUCYTHASASNGLY 360  
DB 1364 ACAAAGCGAACACAGTCACTTGTGTGTCTCTCCAGACTATGACATGATTAATGAC 1423  
QY 361 ILEMETILELATRPASNGLYILEGLUARGLEUARGALAGLYLEUHSISASP 380  
DB 1424 ATTAATGATTTCA----- 1435  
QY 381 ILEGUGLYILEARGTYRGJUPROLYSCYSEPROLEUGLYVALASPILESERLYGLUVAL 400

DB 1436 -----TGATGCTCTTGAGAGTAGACATTCATAAGAGTT 1471  
QY 401 GIVIGLUASERILELYSVALPROGLINLEUYSMETGLUILE 414  
DB 1472 GGAAGCTTCATTAAGTACCAATTAATAATGAGATA 1513  
RESULT 8  
ADQ24627  
ID ADQ24627 standard; DNA; 2890 BP.  
XX  
AC ADQ24627;  
XX  
XX 26-AUG-2004 (first entry)  
DT  
XX  
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 7447.  
XX  
KM soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
KM ds.  
XX  
OS Homo sapiens.  
XX  
EN WO2004048938-A2.  
XX  
PD 10-JUN-2004.  
XX  
XX 26-NOV-2003; 2003WO-US038193.  
PF  
XX 26-NOV-2002; 2002US-0429739P.  
PR  
XX (PROT-) PROTEIN DESIGN LABS INC.  
PA  
XX Aziz N, Ginsburg WM, Zlotnick A;  
PI  
XX WPI; 2004-441208/41.  
DR  
XX  
XX  
PT Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.  
XX  
XX Example 2; SEQ ID NO 7447; 210bp; English.  
PS  
XX  
XX The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cyostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.  
XX  
XX Sequence 2890 BP; 869 A; 609 C; 611 G; 789 T; 0 U; 12 Other;  
SQ  
Alignment Scores:  
Pred. No.: 3.2e-193 Length: 2890  
Score: 1944.00 Matches: 386  
Percent Similarity: 93.24% Conservative: 0  
Best Local Similarity: 93.24% Mismatches: 4  
Query Match: 91.48% Indels: 24  
DB: 12 Gaps: 1  
US-10-649-273-2 (1-414) x ADQ24627 (1-2890)  
QY 1 MetleuileuthrlyserthrAlaGlyValPhepelysProSerlyAsrGlyValYr 20  
DB 1001 ATGCTAATCTTGACATACAGTGCAGAGCTTTTAAACATCAAAAAAGAAATTTAT 1060

```

QY 21 GIUPLLEUARGSERPHEASNPHEHISPROGLYTHREUPHEUHSIYLSYIIIEVALLEU 40
DB 1061 GAATTTTAAAGAGTTTAAATTTTCAATCCGGAACACTATTCTTCAATAAATAAGTATG 1120
QY 41 GYIIIEGLUThrSerCySapPheThrAlaAlaValValAspGluThrGlyAsnVal 60
DB 1121 GGAATTGAAACTGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1180
QY 61 LEUGLYGLUAlaIleHisSerGlnThrGluValHisIleuLeuThrGlyIleValPro 80
DB 1181 TTGGGAGAAACAATACATCCCAACTGAAGTTCATTATAAACAAGTGGATTTGTTCT 1240
QY 81 PROAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
DB 1241 CCAGCAGCTCAACAGCTTCAAGAAATAATTCACAGAAATAGTACAGAAAGCTCTTCT 1300
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLeuProGlyLeu 120
DB 1301 GCCAGTGGAGTCTCTCAAGTGAACCTTCAGCAATTCACATMAAACAGAGACTT 1360
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLeuLys 140
DB 1361 GCTTTAAGCTGGAGAGTGGCTTATATCTTACAGCTTGAGTGGACGTTAAAAAG 1420
QY 141 PROpheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
DB 1421 CCATTTCATCCCATTCATCATATGAGAGCTTCATCTATTTAGTTGACCAATMAA 1480
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
DB 1481 GTAGAAATTCCTTTTATGATTTCTGAGATTCGAGTCACTGCTGGTGGACATTA 1540
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
DB 1541 CAAGAGATTCAGATTTCTGCTTCTTGAAAGCTTTTGACATAGACACAGGAGATG 1600
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220
DB 1601 CTTGACAAAGTGGCAAGAGACTTCCCTTTAATAAACAATCAGAGTCTCCACCATAGT 1660
QY 221 GlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
DB 1661 GGTGGGAAAGCCATAGAACATTTGGCCAAACAAGAAATGATTTGATTTGACATCAA 1720
QY 241 PROProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
DB 1721 CTTCCCTTCATCATGCTAATAAATGTGATTTCTCTTTAATGAGACTTCAACAGTTACT 1780
QY 261 AspLysIleIleMetLysLysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSer 280
DB 1781 GATTAATAATAATAAGAAAAAGAAAAAGAAAGGATTTAGAAAGGGGCAAAATCTGTCT 1840
QY 281 SerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeuValLysArg 300
DB 1841 TCAGCAGCAGACATCTCTGCCACAGTACAGACACACATGCAATGCTCTTGCAAAAGA 1900
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320
DB 1901 AACACATCGGGCTAATCTGTTTGTGAAGCAGAGACTTGTACCTCAAAATAATGACGTA 1960
QY 321 LeuValAlaSerGlyValAlaAlaSerAspPheThrIleArgArgAlaLeuGluIleLeu 340
DB 1961 CTGGTTGATCTGGTGGTGGTGGAGTAACTTCTGTAATCCGAGAGCTCTGGAAATTTTA 2020
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgArgLeuCysThrAspAsnGly 360
DB 2021 ACAAAAGCAACACAGGACCTTTGTTGTGCTCTCTCCAGACATGATCACTGATTAATGCG 2080
QY 361 IleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
DB 2081 AATTAATGATGCA----- 2092
QY 381 IleGluGlyIleArgGlyGluProLysCysProLeuGlyValAspLysSerLysGluVal 400

```

```

DB 2093 -----TGATGTCCTCTTGAGTGAAGCATATCAAAAGAACTT 2128
QY 401 GYGLUAlaSerIleLysValProGlnLeuLysMetGluIle 414
DB 2129 GAGAGACTTCCATTAATAAGTACACATTAATAAATGAGATA 2170

RESULT 9
ABX70950
ID ABX70950 standard; cDNA; 1416 BP.
AC ABX70950;
XX
XX 05-MAR-2003 (first entry)
DE Novel human cDNA sequence #175.
XX
XX Human; gene; ss; nervous system disorder; peripheral neuropathy;
XX Huntington's disease; amyotrophic lateral sclerosis; haemophilia;
XX neurodegenerative disease; Parkinson's disease; Alzheimer's disease;
XX autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;
XX insulin-dependent diabetes mellitus; anaemia; thrombocytopenia; wound;
XX ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;
XX fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;
XX coagulation disorder; cancer; tumour; inflammatory disease; septic shock;
XX Crohn's disease; anaphylaxis; proliferation; chemotactic;
XX differentiation; stem cell growth factor; haematopoiesis; chemokinetic;
XX haemostatic; antiinflammatory; expressed sequence tag; EST.
XX
XX Homo sapiens.
OS
XX MO20281731-A2.
XX
XX 17-OCT-2002.
XX
XX 29-JAN-2002; 2002MO-US001222.
XX
XX 30-JAN-2001; 2001US-00774528.
XX
XX (HXSE-) HXSE0 INC.
XX (GOOD/) GOODRICH R W.
XX
XX Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Dermanac RT;
XX WPI; 2003-058563/05.
XX
XX Novel polypeptide useful for treating neurodegenerative diseases, myeloid
XX or lymphoid cell disorders, bone disorders, mechanical and traumatic
XX disorders, coagulation disorders, and inflammatory diseases.
XX
XX Claim 1; Page: 612pp; English.
XX
XX This invention relates to the cDNA sequences encoding an isolated novel
XX human polypeptide. The protein encoded by the nucleic acid of the
XX invention is useful for treating central and peripheral nervous system
XX diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic
XX lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,
XX Alzheimer's disease); autoimmune disease (e.g. systemic lupus
XX erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)
XX myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopenia)
XX wounds, ulcers, burns; bone disorders (e.g. osteoporosis,
XX osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head
XX trauma); lung or liver fibrosis; reperfusion injury in various tissues;
XX bacterial, viral or fungal infections; allergic conditions such as
XX allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);
XX cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's
XX disease, anaphylaxis). The protein may be used to inhibit the growth,
XX infection or function of infectious agents such as bacteria, fungi,
XX viruses, or to effect bodily characteristics, biohythms or circadian
XX cycles of rhythms. The protein may also have
XX proliferation/differentiation, stem cell growth factor, haematopoiesis
XX regulation, immune stimulating or suppressing, chemotactic/chemokinetic,

```



CC haemostatic and thrombolytic, receptor/ligand, and antiinflammatory  
 CC activities. The cDNA sequences of the invention are useful for expressing  
 CC recombinant protein for analysis. The present sequence represents a novel  
 CC human cDNA sequence of the invention, this sequence is an expressed  
 CC sequence tag (EST) and was identified using subtractive hybridisation  
 XX  
 SQ Sequence 1416 BP, 441 A, 280 C, 273 G, 422 T, 0 U, 0 Other;

## Alignment Scores:

| Pred. No.:             | Length: | 1416            |
|------------------------|---------|-----------------|
| Score:                 | 1747.00 | Matches: 340    |
| Percent Similarity:    | 100.00% | Conservative: 1 |
| Best Local Similarity: | 99.71%  | Mismatches: 0   |
| Query Match:           | 82.21%  | Indels: 0       |
| DB:                    | 8       | Gaps: 0         |

US-10-649-273-2 (1-414) x ABX70950 (1-1416)

QY 74 LysThrGlyGlyIleValProProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArg 93  
 Db 280 AGAACAGGAGGAGTGTCTCCAGCAGCTCAACAGCTTACAGAGAAATATTCACAGCA 339  
 QY 94 IleValGlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAla 113  
 Db 340 ATAGTACAGAGAGCTCTTCTGCGAGTGAGTCTTCAAGTGAACCTTCAGCAATTTGCA 399  
 QY 114 ThrThrIleLeuProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGln 133  
 Db 400 ACTACATATAAACAGAGACTTGTCTTAAGCTGGAGCTGGCTTATCATTTAGCTTACAG 459  
 QY 134 LeuValGlyGlnLeuValSerProPheIleProIleHisIleMetGluAlaIleAlaLeu 153  
 Db 460 CTGCTGAGCAGGATTAATAAAGCCATTCATTCATCATATGAGAGCTCAGCACTT 519  
 QY 154 ThrIleArgLeuThrAsnIleValGlnPheProPheLeuValLeuLeuIleSerGlyGly 173  
 Db 520 ACTATTAGGTGACCATATAAGATATTCCTTTTATGATCTTTGATTTCTGAGAGT 579  
 QY 174 HisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyValSerLeu 193  
 Db 580 CACTGCTGTGGCATTTAGTTCAAGAGGTTTCAGATTTTCTGCTTGGAAAGCTTTTG 639  
 QY 194 AspIleAlaProGlyAspMetLeuAspIleValAlaArgArgLeuSerLeuIleValHis 213  
 Db 640 GACATGACCCAGCTGACATGCTTGAACAAGTGCACAGACCTTTCTTAATATAACAT 699  
 QY 214 ProGluCysSerThrMetSerGlyGlyValAlaIleGlnHisLeuAlaValSerGlnIleAsn 233  
 Db 700 CCAGAGTGCCTCCACCATGAGTGGTGGAAAGCCATAGACATTTGGCCAAACAGGAAT 759  
 QY 234 ArgPheHisPheAspIleValSerProLeuHisIleAlaValAsnCysAspPheSerPhe 253  
 Db 760 AGATTTCAATTTGACATCAAAACCTCCCTGATCATCTAATAAATGATTTTCTTTT 819  
 QY 254 ThrGlyLeuGlnHisValIleThrAspIleIleMetCysValSerGlnGluGlyIle 273  
 Db 820 ACTGGACTTAACACGTTACTGATTAATAATATGAAAAAGAAAAAGAGATATTT 879  
 QY 274 GluValGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaIleValGlnHisThrMet 293  
 Db 880 GAGAGAGGGCAATCTGCTTCCAGCAGACATTCCTGCCACAGTACAGACACATG 939  
 QY 294 AlaCysHisLeuValIleValArgThrHisArgAlaIleLeuPheCysValSerGlnArgAspLeu 313  
 Db 940 GCATGTCATCTTGGAAAGACATCGGCTATTTGTTTGAAGCAGAGACTTG 999  
 QY 314 LeuProGlnAsnAlaValLeuValAlaSerGlyValAlaIleAspAsnPheTyrIle 333  
 Db 1000 TTACTCTCAAAATATATGAGTACTGTGTGATCTGTGTGCGAAGTAACCTTATATATTC 1059  
 QY 334 ArgArgAlaLeuGlnIleLeuThrAspAlaThrGlnCysThrLeuLeuCysProProPro 353  
 Db 1060 CGCAGAGCTCGAAATTTTAAACAAACGACACAGTGCACCTTGTGTGTCTCTCTCC 1119

QY 354 ArgLeuCysThrAspAsnGlyIleMetIleAlaIleProAlaGlyIleGlnArgLeuAla 373  
 Db 1120 AGACTATGCACTGATTAAGGATTAATGATTCATGAGATGATTTGAAAGACTAGTCT 1179  
 QY 374 GlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGlnProLeuCysProLeuGly 393  
 Db 1180 GCGTTGGCATTTTACATGACATGAGCAGTCCGCTATGAAACCAATATGCTCTTGA 1239  
 QY 394 ValAspIleSerGlyGluValGlyGluAlaSerIleValProGlnLeuValSerMetGlu 413  
 Db 1240 GTAGACATATCAAAAGAAAGTTGAGAGACCTTCATTAAGTACCAATTAATAATGAG 1299  
 QY 414 Ile 414  
 Db 1300 ATA 1302

## RESULT 10

AB075508 standard; DNA, 1572 BP.

AB075508;

07-NOV-2002 (first entry)

Murine sialoglycoprotease-like gene sequence SEQ ID NO:7.

Murine; mouse; protease; calcium activated neutral protease type 5;  
 KRN5; trypsinase 4; sialoglycoprotease; enzyme; genetic disease;  
 KRN5; neurological; neuropsychological; psychotic illness; transgenic animal;  
 gene; ds.

Mus musculus.

W0200245491-A2.

13-JUN-2002.

05-DEC-2001; 2001WO-US046405.

06-DEC-2000; 2000US-0251803P.

06-DEC-2000; 2000US-0251820P.

13-DEC-2000; 2000US-0255971P.

(DELTA-) DELTACEN INC.

Allen KD, Leviten MW;

WPI; 2002-657389/70.

Novel transgenic animal, comprising a disruption in protease target gene,

is useful for identifying agents that ameliorates a phenotype associated

with a disruption in a protease target gene.

Example 3; Fig 7; 62pp; English.

The present invention describes a non-human transgenic animal (I) comprising a disruption in a protease target gene (PG) selected from calcium activated neutral protease type 5 (CAPN5) gene, trypsinase 4 gene and sialoglycoprotease-like gene. Also described is a targeting construct (II), comprising a first polynucleotide sequence homologous to at least a first portion of PG, a second polynucleotide sequence homologous to at least a second portion of PG and a selectable marker. (II) is useful for producing a transgenic mouse comprising a disruption in a protease target gene, by introducing (II) into a cell, introducing the cell into a blastocyst, implanting the resulting blastocyst into a pseudopregnant mouse, where the pseudopregnant mouse gives birth to a chimeric mouse, and breeding the chimeric mouse to produce the transgenic mouse. (I) is useful for identifying an agent that modulates the expression or function of a protease target gene, by administering an agent to (I) and determining whether the expression or function of the disrupted protease target gene in (I) is modulated. (I) is also useful for testing the efficacy of proposed genetic and pharmacological therapies for human



CC genetic diseases, such as neurological, neuropsychological or psychotic  
 CC illness. The present sequence represents murine staloglycoprotease-like  
 CC gene sequence, which is used in an example from the present invention  
 XX  
 SQ Sequence 1572 BP; 459 A; 337 C; 340 G; 429 T; 0 U; 7 Other;

## Alignment Scores:

| Pred. No.:             | 3,766-152 | Length:       | 1572 |
|------------------------|-----------|---------------|------|
| Score:                 | 1549.00   | Matches:      | 326  |
| Percent Similarity:    | 84.50%    | Conservative: | 23   |
| Best Local Similarity: | 78.93%    | Mismatches:   | 60   |
| Query Match:           | 72.89%    | Indels:       | 7    |
|                        |           | Gaps:         | 2    |

US-10-649-273-2 (1-414) x ABQ75508 (1-1572)

```

QY 1 MetLeuIleLeuThrIleThrAlaGlyValPhePheLeuProSerIleValTyr 20
Db 5 ATGCTAATGTTAAAGAGACAGACAGAGCTATTCCCAAGCCCAAGAGTAAAGTTAT 64
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValLeu 40
Db 65 GGATTTTAAAGAGATTAGTTCATCCAGAACTCTCTTGTCTATAACTGGTCTG 124
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60
Db 125 GGATTTGAAACACAGCTGTATGACACAGAGCGGCTGTGTGATGAAATGGGAATGTG 184
QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuIleValPro 80
Db 185 CTGGGGGAGACACTGCATCCCAACTCAGGTCATCTGAAAACAGTGGGATTTGTCCT 244
QY 81 ProAlaIleGlnGlnIleuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
Db 245 CCAGTAGCTCAACACTTCACAGAGAAATATTCACGAATAGAGAGAACTCTTCT 304
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrIleValProGlyLeu 120
Db 305 GGCTTAGAATATCAACCCAGCGATCTCTCAGCAATTCACATCAACCGGAGCTG 364
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnIleuValGlyIleuValys 140
Db 365 GCCCTAAGTTGGGAGATTGGCTTATCTTTCCTTACAGCTAGTAATCAGTTTAAARS 424
QY 141 ProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeuThrAspIle 160
Db 425 CCATTTCAT-CCGATTATCATCATGAGGCTCAGCGACTGACTATAGGCTCACCAGTAA 483
QY 161 ValGluPheProPheLeuValIleuLeuIleSerGlyValHisCysLeuLeuAlaLeuVal 180
Db 484 GTRAGATTN-CTTTTATTAGTTCTTTGATTCTGGCGGACATCGCTGGKRTTAGTC 542
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyValSerLeuAspIleAlaProGlyAspMet 200
Db 543 CAAGGATTTCGATTTCTGCTCTTGGGAGAGCTTTGGACATAGACACAGGAGCATG 602
QY 201 LeuAspIleValAlaAlaArgArgLeuSerLeuIleValHisProGlyCysSerThrMetSer 220
Db 603 CTTGACAAAGGTGCAAGAAAGACTTCTTTAATCAAAACATCCAGAAATGTTCTTACATAGCT 662
QY 221 GlyGlyValAlaIleGlnHisLeuAlaValysGlnGlyAsnArgPheHisPheAspIle 240
Db 663 GGTGAAAAGCTTATGAAACGTTGGCCAAAGACGAAATAGATTCCATTTACTATGAAT 722
QY 241 ProProLeuHisHisAlaValysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
Db 723 CCACCTATGACAGAAATCTAAGATTCGATTTCTTTTCAAGGAGCTTCAACATATTACT 782
QY 261 AspIleValIleIleMetIleValysGlyValGluGlyIleGlnIleValysGlnIleLeuSer 280
Db 783 GATTAAGCTAATATACACAAAGGAAAGAAAGAGGCAATGAGAGGGGCAAAATCTGTCA 842
QY 281 SerAlaIleAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValValArg 300

```

```

Db 843 TCAGCTGACAGACATTGCTGCTGCGGTACAGATGCCAAAGGTCACCTTGCAAAAGA 902
QY 301 ThrHisArgAlaIleLeuPheCysValysGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320
Db 903 ACACATGCGCTATTCTGTTTTCAGACAGAGAAATTTGCTCTCTCCAGCTAACGCACTA 962
QY 321 LeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArgAlaLeuGluIleLeu 340
Db 963 TTAGTTTATCTGGAGGTGTGGCAAGTAACCTGACATCCGAAAGACATTTGAAATTGTC 1022
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360
Db 1023 GCAATGCAACGCATGACAGCGTGTGTGTCTCCT-TCAGACCTGTGACAGCAATGGG 1081
QY 361 IleMetIleAlaThrPheAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeu-HisAs 380
Db 1082 CATATGATTCATGCAATGAAATGGAATGGAATGATTAAGTGTGCTGCTGCTTTTACCA 1141
QY 380 P-IleGluGlyIleArgTyrGluProIleCysProLeuGlyVal---AspIleSerIle 399
Db 1142 TGATAGAGACATTGCGTTATTAAACCAATGTCTCTTTGATGAGGATTAATCCGAG 1201
QY 399 IuValGlyGluAlaSerIleValValProGlnLeu 410
Db 1202 AAGTTGGCAGAA---AGCTTGCCCATTAATAAGTTA 1233

```

RESULT 11  
 ID ABS76639 standard; DNA, 1526 BP.  
 AC ABS76639;  
 XX  
 DT 11-DEC-2002 (first entry)  
 XX  
 DE DNA encoding novel human metalloprotease MP1 fragment #1.  
 XX  
 KW Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;  
 KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
 KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
 KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
 KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
 KW liver disease; renal disease; immune disorder; rheumatoid arthritis;  
 KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
 KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
 KW neurological disorder; gene; ds.  
 KW  
 OS Homo sapiens.  
 XX  
 PN WO200272751-A2.  
 XX  
 PD 19-SEP-2002.  
 XX  
 PF 05-FEB-2002; 2002WO-US003353.  
 XX  
 PR 05-FEB-2001; 2001US-0266518P.  
 PR 10-APR-2001; 2001US-0282814P.  
 XX  
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S;  
 XX WPI; 2002-723329/78.  
 DR P-PSDB; ABG96487.  
 XX  
 PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
 PT treating, or ameliorating diseases associated with aberrant  
 PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and  
 PT neurological disorders.  
 XX  
 PS Disclosure; Page 462-463; 473pp; English.



hepatocytic, virucide, haemostatic, anti-HIV, antithyroid, thyromimetic, chemotherapeutic, antibacterial, fungicide, antiparasitic, anticoagulant, thrombolytic, anticoagulant, anorectic, vasotropic, and antiviral. The novel D11P polynucleotides polypeptide can be used in gene therapy and protein replacement therapy. The d11p polynucleotides or D11P polypeptides are useful for diagnosing, preventing or treating diseases associated with the expression of human molecules. In particular, these diseases include cancers (e.g. adenocarcinoma, leukemia, melanoma, Brain cancer, breast cancer, cervix cancer, bone cancer, liver cancer, lung cancer) or other cell proliferative disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, polycythemia vera, primary thrombocytopenia), autoimmune/inflammatory disorders (e.g. AIDS, Addison's disease, thyroiditis, Crohn's disease, Graves' disease, Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid arthritis, scleroderma, systemic lupus erythematosus), infections (e.g. viral, bacterial, fungal or parasitic infection), developmental disorders (e.g. Cushing's syndrome or epilepsy), endocrine disorders (e.g. chromoblast, hypopituitarism, hypoadrenalism, gigantism, goiter) metabolic disorders (e.g. hypercholesterolaemia, hypoglycaemia, diabetes, hyperlipidaemia, obesity), neurological disorders (e.g. ischaemic cerebrovascular disease, stroke, Alzheimer's disease, Pick's disease, Huntington's disease, Parkinson's disease, Creutzfeldt-Jakob disease, anxiety, schizophrenia), gastrointestinal disorders (e.g. ulcers), transport disorders (e.g. akinesia or multidrug resistance), or connective tissue disorders (e.g. Paget's disease or rickets). This polynucleotide sequence represents one of the human d11p DNA sequences of the invention.

Sequence 3358 BP; 1105 A; 577 C; 601 G; 1075 T; 0 U; 0 Other;

**Alignment Scores:**

|                        |          |               |      |
|------------------------|----------|---------------|------|
| Pred. No.:             | 3,39e-94 | Length:       | 3358 |
| Score:                 | 1001.50  | Matches:      | 250  |
| Percent Similarity:    | 44.82%   | Conservative: | 1    |
| Best Local Similarity: | 44.64%   | Mismatches:   | 7    |
| Query Match:           | 47.13%   | Indels:       | 307  |
| GB:                    | 10       | Gaps:         | 2    |

|    |      |   |      |
|----|------|---|------|
| QY | 157  | LeuThrAsnLysValGluPheProPheLeuValLeuLeuLisSerGlyGlyYHIsCyLeu      | 176  |
| Db | 3067 | GTGACCAATTAAGAGAGAAATTTCTTTTAACTCT-TTGAATTTCTGAGAGTCACCTGCTG      | 3009 |
| QY | 177  | LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlnGlyLysSerLeuAspLLeaIa      | 196  |
| Db | 3008 | TGGCATTAGTTCACAAGAGTTTCAGATTTTCTGCTTCTTGAAAAGTC-TTGACACATACCA     | 2950 |
| QY | 197  | ProGlyAspMet-----   | 200  |
| Db | 2949 | CCAGGTGACATGCGCTTGACAAAGTAATTAGAATAATTCTCCATCTTTTGTATGT           | 2890 |
| QY | 200  | -----   | 200  |
| Db | 2889 | TGTCATTTCAACTAAGTAGACAATAGATGTGCTACACATTTCACCTAATATTTCGAA         | 2830 |
| QY | 200  | -----   | 200  |
| Db | 2829 | TTTTATCTTAGTAATACTGCACACAAAAATTCACATATGTGTGAGAAAAATAGAAAACAGTA    | 2770 |
| QY | 200  | -----   | 200  |
| Db | 2769 | GTACAGCAATTTATTAATTTCTTAGCCTTCTTAATTAATAATGTATAGAGTTCAATCTGTA     | 2710 |
| QY | 200  | -----   | 200  |
| Db | 2709 | CATTAAGGCTGAAATAGTTTGACAGATACAGTTATGATTTTGCAAATATATGATGTGAA       | 2650 |
| QY | 200  | -----   | 200  |
| Db | 2649 | AGAACTGCTTCGTAATACTAACATACATCTGCAGAAAAAGGTAAAAATAGAGAAATATATATGCA | 2590 |
| QY | 200  | -----   | 200  |

[illegible]

DB 1512 GCATTTACATGACATAGAAAGCATCCGCTATGAAACCAAAATGCTCTTGAGTAGACA 1453  
QY 396 IserterysgluValuGluAlaSerileysValProgluInleuYemecGluile 414  
DB 1452 TATCAAAAGAAAGTTGGAGAACTTCCTAAAGAACCAATTTAAATGAGATA 1397  
RESULT 13  
ABL24633  
ID ABL24633 standard; DNA; 1601 BP.  
AC ABL24633;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 25372.  
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 25372.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX  
XX Drosophila melanogaster.  
OS  
PN WO200171042-A2.  
XX  
XX 27-SEP-2001.  
PD 23-MAR-2001; 2001WO-US009231.  
PF 23-MAR-2000; 2000US-0191637P.  
XX 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-656860/75.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
XX Claim 1; SEQ ID NO 25372; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signaling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB16511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins (AB57737-  
CC AB872072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 1601 BP; 344 A; 477 C; 441 G; 339 T; 0 U; 0 Other;  
SQ  
Alignment Scores:  
Pred. No.: 1.51e-64 Length: 1601  
Score: 714.50 Matches: 156  
Percent Similarity: 58.22% Conservative: 60  
Best Local Similarity: 42.05% Mismatches: 140  
Query Match: 33.62% Indels: 15  
DB: 4 Gaps: 6  
US-10-649-273-2 (1-414) x ABL24633 (1-1601)  
QY 39 ValuGluValuGluThrSerCyaspAspThrAlaAlaValuValuAspGluThrGly 58  
DB 79 GTCTGGGCAATCGACCTCTCGACACAGGCGCATCGCCTGTGACACCAAGGCGC 138  
QY 59 AsnValuGluValuAlaIleHisSerGlnThrGluValuHisIleuYsThrGlyGlyIle 78  
DB 139 CGAGTGATTCGCAATGTCGTGAGTGGCAACAGGAGTTCCACACCCGCTATGAGGCAATT 198

QY 79 ValProAlaAlaGlnGlnIleuHisArgGluAsnIleGlnArgIleValGlnGluAla 98  
DB 199 ATTCGCTCCAGGCGCCAGACCTTCAACCGCCGACATCGCCCTACAGGCTGC 258  
QY 99 LeuSerIleSerGlyValSerProSerAspIleuSerAlaIleAlaThrThrIleYsPro 118  
DB 259 ATGAGAGGGCGCAATGAAAGCCGACCACTGACGCGCAATCGGTACACAGCGCTCC 318  
QY 119 GlyLeuAlaLeuSerIleuGlyValGlyLeuSerPheSerIleuGlnLeuValGlyGlnLeu 138  
DB 319 GCACTGCCCGCTGAGTTTCGTGCGCGCGCTTGACAGCGCACTGCGCCGCTGC 378  
QY 139 LysIleProPheIleProIleHisIleMetGluAlaHisAlaLeuThrThrIleArgLeuThr 158  
DB 379 CAGAAGCCCTGCTCCGCTTACACCAATGAGAGGCGCAGCGCTGCAGCGCCGCTGCA 438  
QY 159 Asn-----LysValGluPheProPheIleuValLeuIleSerGlyGlyHisCysLeu 176  
DB 439 CATCCGAGCAGATCGCTACCCCTCTGCTGTGTGGCCGCGCGCCACTGTGAG 498  
QY 177 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 196  
DB 499 TTGCTGTGCTTAACGCGCCGCTGCTTACGCTTCTTGCGCAACATGATATGCG 558  
QY 197 ProGlyAspMetLeuAspLysValAlaArgArgLeuSerIleuLysHisProGluCys 216  
DB 559 CCGGCGAGCGCTTGAACAAGTCGCGGACATACGCTGACATTCGCGGAGTAC 618  
QY 217 SerThrMetSerGlyLysAlaIleGluHisIleuAlaLys---GlnGlyAsnArgPhe 235  
DB 619 CGTTGTGGAACGAGGAGCGGCGCAATTGAGATGCGCGCCAGCTGCGCAGCATCGCTG 678  
QY 236 HisPheAspIleLysProProIleuHisIleAlaLysAsnCysAspPheSerPheThrGly 255  
DB 679 GCTACGAGTTTCGCTGCGCACTTCCCAAGCGGAACTGCACATTCAGCTCCCGCG 738  
QY 256 LeuGlnHisValThrAspLysIleIleMetLysGlyLysGluGluGlyIleGluLys 275  
DB 739 ATCAAGAACATCTCTTCGCGGCGCATCGAGCGCGGAGCGGAGGAGCACTCGCG 798  
QY 276 GlyIleIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 295  
DB 799 GACGAGATCATCAGACATCAGCGCATCTTTCGCGCGCTGCTGCTGTCAGCGCG 858  
QY 296 HisLeuValLysArgThrHisArgAlaIleLeuPheCys-----LysGlnArgAspLeu 313  
DB 859 CACCTGAGCAACCGGACGAGCGGCAATGAGTACTGCTCTGCGGCGGAGCGAGCTC 918  
QY 314 LeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsnPheTyrlle 333  
DB 919 TTCGCTGACACCGCCGCACTGATCATGTCGCGGCTGAGGCGCAACAGATGCCAT 978  
QY 334 ArgArgAlaLeuGluIleuThrAsnAlaThrGlnCysThrLeuLeuCysArgProPro 353  
DB 979 TACGCCAATCGAATCATCTGCGCGCGAGTATGCTGAGAGAGCTTCGCTCATCGAAG 1038  
QY 354 ArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAla 373  
DB 1039 CGGTACTCTCCGACACAGCGCTCATGATCGCTGCGGCGGCTGAGAGCACTG----- 1092  
QY 374 GlyLeuGlyIleLeuHisAspIleGlu---GlyIleArgTyrgLysProLysCysProLeu 392  
DB 1093 -----CTGCAGATTAAGAGGCGGACGAGCGGCTACGATCAGAC----- 1131  
QY 393 GlyValAspIleSerLysGlyValAlaGlyAla 403  
DB 1132 AGCATTTGATTCAGAGGCGAGCGCGGATTCGCC 1164  
RESULT 14  
AAH15110/C  
ID AAH15110 standard; cDNA; 1385 BP.  
XX  
XX AAH15110;

XX 26-JUN-2001 (first entry)  
 XX Human cDNA sequence SEQ ID NO:13144.  
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 PN EP1074617-A2.  
 PD 07-FEB-2001.  
 PF 28-JUL-2000; 2000EP-00116126.  
 PR 29-JUL-1999; 99JP-00248036.  
 PR 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-00118776.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 09-JUN-2000; 2000JP-00241899.  
 XX (HELI-) HELIX RES INST.  
 PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saico K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX MPI; 2001-318749/34.  
 DR  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 PT length cDNAs defined in the specification, and for the detection and/or  
 PT diagnosis of the abnormality of the proteins encoded by the full-length  
 PT cDNAs.  
 PS Claim 8; SEQ ID NO 13144; 2537pp + Sequence listing; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602 full-  
 CC length cDNAs defined in the specification. Where a primer set comprises:  
 CC (a) an oligo-dr primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 CC  
 XX Sequence 1385 BP; 439 A; 237 C; 256 G; 453 T; 0 U; 0 Other;  
 SQ  
 XX

Alignment Scores:  
 Pred. No.: 3.6e-63 Length: 1385  
 Score: 7.00.50 Matches: 171  
 Percent Similarity: 38.31% Conservative: 1  
 Best Local Similarity: 38.08% Mismatches: 1  
 Query Match: 32.96% Indels: 277  
 DB: 4 Gaps: 1

US-10-649-273-2 (1-414) x AAH15110 (1-1385)  
 QY 74 LysThrGlyValIleValProPheAlaGlnIleuHisArgGluAsnIleGlyArg 93  
 DB 1346 AGAACAGGTGGATGTTCTCCAGCAGCTCAACAGCTTCACAGAGAAATTTCAACGA 1287

QY 94 IleValGlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAla 113  
 DB 1286 ATACTACAGAGAGCTCTTTCGCCAGTGAAGTCTCCAAAGTAGACCTCGACCAATTCA 1227  
 QY 114 ThrThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGln 133  
 DB 1226 ACTACCAATTAACACAGGACTGCTTTAAGCTGGAGAGGCTTATCATTTAGCTTACAG 1167  
 QY 134 LeuValGlyGlnLeuLysLysProPheIleProIleHisIleMetGluAlaHisAlaLeu 153  
 DB 1166 CTGGTAGACAGCTTAAACCAATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1107  
 QY 154 ThrIleArgLeuThrAsnLysValGluPheProPheLeuValLeuLeuIleSerGly 173  
 DB 1106 ACTATTAGCTTGACCAATTAAGTGAATTCCTTTTAGTCTCTTGAATTTCTGGAGGT 1047  
 QY 174 HisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeu 193  
 DB 1046 CACTGCTGTTGGCATTAGTTCAAGGAGTTCAAGATTTTCTGCTTCTTGAAAGTCTTGG 987  
 QY 194 AspIleAlaProGlyAspMetLeuAspLysVal----- 204  
 DB 986 GACATGACACCAAGGTGACATGCTTACAGAGGT-AATTAGAATTAAATTTCTCATTTCTT 928  
 QY 204 ----- 204  
 DB 927 TTTGTATGTTGTCATTTCACTAGTAGCAATAGATGCTACACCATTCACCTTAA 868  
 QY 204 ----- 204  
 DB 867 TATTTGGAATTTATCTAGTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 808  
 QY 204 ----- 204  
 DB 807 AAGAGTAGTACACAATTTATTAATTTCTAGCTTTCTTAATTAATTAATTAATTAATTAAT 748  
 QY 204 ----- 204  
 DB 747 ATCTGTACATTAAGGCTGAATTAAGTTGACATACAGTTAGTATTTGGCAATTAATGT 688  
 QY 204 ----- 204  
 DB 687 ATGTGAAGAAGCGTTCGTAACCTAATCACTAGTCAAAAAAGGTAAATAAGAAATAT 628  
 QY 204 ----- 204  
 DB 627 ATATAGATTAAACATTAAGGACATTAAGATGCAATGACAGATTAATCAACAATTACT 568  
 QY 204 ----- 204  
 DB 567 TACACCAAGAGAGGTCCTCCGACACCCCTTTGTTAGTAATATCAAGAGGCTACT 508  
 QY 204 ----- 204  
 DB 507 GCCATATATAGGAAAAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 448  
 QY 204 ----- 204  
 DB 447 TAGAGATATAGACAAGTCTTATTAATTAAGCTTCATTAAGCAAGTATTTGCAACT 388  
 QY 204 ----- 204  
 DB 387 TCAAGCCATTTTCCAAACAATAGAGCAACAATAGACAGGGCAGTATTTGGCTCT 328  
 QY 204 ----- 204  
 DB 327 TATTTGCGGTCAATCAATAGGAACAGGTTGTCTGCTTACCTGAATATCACTATAGTC 268  
 QY 204 ----- 204  
 DB 267 TATATTTGCCAAGATATAGCATGTTTATTCATTCAGGGGTTTTTTGTTTGTAGTAAT 208

QY 204 ----- 204  
Db 207 TTCAATTATTCCTTGCATCTTTCTTTACAGTATTTATTTATGACTCTAAA 148  
QY 205 ----- 205  
Db 147 AATATGTTCTTGTGATAGTGGCAAGAGACTTTCTTTAATAAACAATCCAGAGTCTCC 88  
QY 218 ThMetSerGlyGlyValAlaIleGluHisLeuAlaIleGlnGlyValAsnArgPheHisPhe 237  
Db 87 ACCATAGTGTGTGGAAAGCCATGACATTTGGCCAAACAGAAATAGATTTCATTTT 28  
QY 238 AspIleLysProPheLeuHisHisAla 246  
Db 27 GACATCAAACTCCCTTGACATCATGCT 1  
RESULT 15  
ID ABL24632/c  
XX ABL24632 standard; DNA; 3656 BP.  
AC ABL24632;  
XX 26-MAR-2002 (first entry)  
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 25369.  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical; gene; ds.  
XX Drosophila melanogaster.  
XX WO200171042-A2.  
XX 27-SBP-2001.  
PD 23-MAR-2001; 2001WO-US009231.  
PF 23-MAR-2000; 2000US-0191637P.  
XX 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li FWD, Myers EW;  
PI WPI; 2001-656860/75.  
DR  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
XX Claim 1; SEQ ID NO 25369; 21bp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins (AB857737-  
CC AB872072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 3656 BP; 1035 A; 829 C; 874 G; 918 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 3.5e-61 Length: 3656  
Score: 687.50 Matches: 157  
Percent Similarity: 55.64% Conservative: 60  
Best Local Similarity: 40.26% Mismatches: 139  
Query Match: 32.35% Indels: 34  
DB: 4 Gaps: 7

US-10-649-273-2 (1-414) x ABL24632 (1-3656)  
QY 39 ValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValAspGluThrGly 58  
Db 2578 GTCTTGCGGCAATCGAGACTCTCTGGACAGACAGCGGCATCGGCATGTGGACACCGGCG 2519  
QY 59 AsnValLeuGlyGluAlaIleHisSerGlnThrGluValHis----- 72  
Db 2518 CGAGTGAATTGCCAATGTCTGAGTCGACACAGAGTTCCACACCCGGTGAAGTCTACA 2459  
QY 73 -----LeuValThr-GlyGlyIleValPr 80  
Db 2458 TCTCCGATGCGGTGCCAATGATGACTGAATGATGACTGAATGATGATGATGATGATGATGAT 2399  
QY 80 oProAlaIleGlnGluHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSe 100  
Db 2398 GCCCAGGCGCCAGACCTTACCGCGCCGCGATGACGTCGCTACACGCGCTGATGGA 2339  
QY 100 rAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLe 120  
Db 2338 GCGCGCGCAATGGAAGCCGACCAACTGACGCGCATCGCGTGACACGCGTCCGCGACT 2279  
QY 120 uAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140  
Db 2278 GCCCTGAGTTTGTGTGTGGCGGTGCTTGCACGCGACCTGCGCCGTGCTGACAGA 2219  
QY 140 sProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsn-- 159  
Db 2218 GCCCTGTGCGCGTTCACACATGAGAGCGGACCGGTCGACGCGCCGATGGAACATTC 2159  
QY 160 ----LysValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAl 178  
Db 2158 GGAGCAGATCGGCTACCTCTCTGCTGTGCGCAGCGCGGCGCATGTCAGTTGCT 2099  
QY 178 aLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGly 198  
Db 2098 GTGGCTTAACGCGCCGCGCTTACCTTGTGGCAACATGATGATGATGATGATGATGATGATGAT 2039  
QY 198 yAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThr 218  
Db 2038 CGAGGCTTTGACAGATCGGTGCGGCTGACGCTGACATTCGCGGAGTACCGCTT 1979  
QY 218 rMetSerGlyGlyLysAlaIleGluHisLeuAlaLys--GlnGlyAsnArgPheHisPhe 237  
Db 1978 GTGACAGAGAGACCGGCATTTGACATCCGCCACGCGGCGACGACGATCCGCTGCTCA 1919  
QY 237 eAspIleLysProPheLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGly 257  
Db 1918 CGAGTTTCCGCGTCCACTTGGCCAGCAGAGGAACTGCAACTTCAGCTTCCGCGGATCA 1859  
QY 257 nHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLysGlyGly 277  
Db 1858 GAACAACTCTCTCCGCGCATTCGACGCGGCGAGCGGCGAGCGGACGACCTCCGACCG 1799  
QY 277 nIleLeuSerSerAlaAlaAspIleAlaThrValGlnHisThrMetAlaCysHisIle 297  
Db 1798 AGTCAATGACCACTTACCGGCACTTCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1739  
QY 297 uValLysArgThrHisArgAlaIleLeuPheCys-----LysGlnArgAspLeuLeuPr 315  
Db 1738 GATGACCGGACGACGCGGCGGATTTGATGATGCTCTCTGCGGCGACGACGACCTTCTGCG 1679  
QY 315 oGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgAr 335  
Db 1678 TGACACCCCGCCACACTGTCATGTCGCGGCGTGGCCACACAGATGACATATACGC 1619  
QY 335 gAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLe 355  
Db 1618 CAACATCCAAATCTTCCGCGCGCTAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1559  
QY 355 uCysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLe 375  
Db 1558 CTGCTCCGACCAAGCGCGTCAATGATGCTGCGGACGCGCTTGACACATG----- 1510

|    |      |   |      |
|----|------|---|------|
| Qy | 375  | uglylleuHiaaplleglu---GlylleArgTyrGluProLysCysProLeuGlyVa | 394  |
|    |      |   |      |
| Db | 1509 | -----CTGCAGGATTAAGAGGCCACAGCGCGCTACGACTACGAC-----AGCAT    | 1466 |
|    |      |   |      |
| Qy | 394  | laaplleserlyesgluValGlyGluAa                              | 403  |
|    |      |   |      |
| Db | 1465 | TGATATTCAGGGCAGCGCGGATTCGCC                               | 1438 |

Search completed: June 16, 2005, 20:14:57  
 Job time : 636.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 16, 2005, 23:18:53 ; Search time 2812.5 Seconds

(without alignments)  
913.751 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125

Sequence: 1 MLILKTAGVFFKPSKRVY.....DISKEVGASIKYQQLMEI 414

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 6054689 segs, 3103772919 residues

Total number of hits satisfying chosen parameters: 12109378

Minimum DB seg length: 0  
Maximum DB seg length: 2000000000.

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool.p/US10649273/runat.15062005.111418.6138/app.query.fasta.1.1429  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62  
-TRANS=human40.cdi -LIST=45 -DOCCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MNTLEN=0  
-MAXLEN=2000000000 -USER=US10649273 @CGN 1 1 1034 @runat.15062005.111418.6138  
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSELOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FAPOP=6 -FAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications NA:\*

1: /cgn2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/prodata/1/pubpna/PCR\_NEW\_PUB.seq.\*  
3: /cgn2\_6/prodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/prodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
5: /cgn2\_6/prodata/1/pubpna/US06\_PUBCOMB.seq.\*  
6: /cgn2\_6/prodata/1/pubpna/PCRTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/prodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq.\*  
10: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/prodata/1/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/prodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq.\*  
14: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq.\*  
15: /cgn2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq.\*  
17: /cgn2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
18: /cgn2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
19: /cgn2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
20: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq.\*  
21: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq.\*  
22: /cgn2\_6/prodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
23: /cgn2\_6/prodata/1/pubpna/US11\_PUBCOMB.seq.\*  
24: /cgn2\_6/prodata/1/pubpna/US11\_NEW\_PUB.seq.\*  
25: /cgn2\_6/prodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
26: /cgn2\_6/prodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Query Length | DB ID | Description          |
|------------|--------|-------------|--------------|-------|----------------------|
| 1          | 2125   | 100.0       | 2197         | 14    | US-10-067-443-1      |
| 2          | 2125   | 100.0       | 2197         | 18    | US-10-649-273-1      |
| 3          | 2125   | 100.0       | 2197         | 18    | US-10-651-722-1      |
| 4          | 2125   | 100.0       | 2572         | 21    | US-10-480-988-36     |
| 5          | 2090.5 | 98.4        | 1387         | 14    | US-10-067-443-21     |
| 6          | 2090.5 | 98.4        | 1387         | 18    | US-10-649-273-21     |
| 7          | 2090.5 | 98.4        | 1387         | 18    | US-10-651-722-21     |
| 8          | 2090.5 | 98.4        | 1387         | 21    | US-10-887-553A-1047  |
| 9          | 2088   | 98.3        | 1245         | 14    | US-10-012-140-6      |
| 10         | 2088   | 98.3        | 1820         | 14    | US-10-012-140-4      |
| 11         | 1944   | 91.5        | 2208         | 17    | US-10-094-749-400    |
| 12         | 1944   | 91.5        | 2890         | 20    | US-10-723-860-7447   |
| 13         | 1747   | 82.2        | 1416         | 17    | US-10-120-988-177    |
| 14         | 1385   | 65.2        | 1526         | 14    | US-10-067-443-23     |
| 15         | 1385   | 65.2        | 1526         | 18    | US-10-649-273-23     |
| 16         | 1385   | 65.2        | 1526         | 18    | US-10-651-722-23     |
| 17         | 1186.5 | 55.8        | 14364        | 14    | US-10-067-443-20     |
| 18         | 1186.5 | 55.8        | 14364        | 18    | US-10-649-273-20     |
| 19         | 1186.5 | 55.8        | 14364        | 18    | US-10-651-722-20     |
| 20         | 682    | 32.1        | 1917         | 18    | US-10-424-599-66417  |
| 21         | 578    | 27.2        | 1628         | 19    | US-10-437-963-11249  |
| 22         | 531.5  | 25.0        | 4360         | 14    | US-10-081-051-2      |
| 23         | 524    | 24.7        | 1146         | 17    | US-10-282-122A-14674 |
| 24         | 502    | 23.6        | 1044         | 17    | US-10-282-122A-26972 |
| 25         | 502    | 23.6        | 94750        | 18    | US-10-672-787-38     |
| 26         | 494.5  | 23.3        | 1032         | 17    | US-10-282-122A-31043 |
| 27         | 492.5  | 23.2        | 1026         | 9     | US-09-815-242-7701   |
| 28         | 492.5  | 23.2        | 1196         | 17    | US-10-282-122A-30016 |
| 29         | 488.5  | 23.0        | 1124         | 20    | US-10-425-115-167432 |
| 30         | 484.5  | 22.8        | 1014         | 9     | US-09-815-242-6207   |
| 31         | 484.5  | 22.8        | 1014         | 17    | US-10-282-122A-20511 |
| 32         | 484.5  | 22.8        | 1014         | 17    | US-10-282-122A-39301 |
| 33         | 483.5  | 22.8        | 1014         | 17    | US-10-282-122A-41977 |
| 34         | 482.5  | 22.7        | 1020         | 17    | US-10-282-122A-32254 |
| 35         | 481.5  | 22.7        | 1014         | 9     | US-09-815-242-9686   |
| 36         | 475.5  | 22.4        | 1029         | 17    | US-09-815-242-6946   |
| 37         | 475.5  | 22.4        | 1029         | 17    | US-10-282-122A-22020 |
| 38         | 475.5  | 22.4        | 1830121      | 17    | US-10-329-670-1      |
| 39         | 475.5  | 22.4        | 1830121      | 20    | US-10-158-865-1      |
| 40         | 469    | 22.1        | 1023         | 17    | US-10-282-122A-15870 |
| 41         | 468    | 22.0        | 371          | 19    | US-10-430-201-3118   |
| 42         | 468    | 22.0        | 371          | 19    | US-10-430-201-3119   |
| 43         | 468    | 22.0        | 1014         | 17    | US-10-282-122A-21633 |
| 44         | 467.5  | 22.0        | 1007         | 17    | US-10-282-122A-19220 |
| 45         | 465.5  | 21.9        | 1065         | 17    | US-10-282-122A-29584 |

#### ALIGNMENTS

RESULT 1  
US-10-067-443-1  
Sequence 1, Application US/10067443  
Publication No. US20030082782A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
FILE REFERENCE: D0073 NP  
CURRENT APPLICATION NUMBER: US/10/067,443  
CURRENT FILING DATE: 2002-02-05  
PRIOR APPLICATION NUMBER: US 60/266,519  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: US 60/282,814  
PRIOR FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71



```
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
US-10-649-443-1

Alignment Scores:
Pred. No.: 2,07e-247 Length: 2197
Score: 2125.00 Matches: 414
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-10-649-273-2 (1-414) x US-10-649-443-1 (1-2197)
QY 1 MetLeuIleLeuThrIleThrsThrAlaGlyValPhePheIysProSerIysArgLysValTyr 20
DB 231 ATGCTTAATCTTGACTTAAGACTGCAAGAGTTTCTTTTAAACCATCAAAAGAAAGTTTAT 290
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValLeu 40
DB 291 GAATTTTAAAGAGTTTATTTATTTTCATCCCTGAAACACTATTTCTCTTAATAATAGTATG 350
QY 41 GlyIleGlyThrSerCysAspAspThrAlaAlaAlaValAspGluThrGlyAsnVal 60
DB 351 GGAATTTGAAACTGATGTTGTGATGATACAGCAGCTGCTGTGTGTGATGAAACTGGAATGTG 410
QY 61 LeuGlyValAlaIleHisSerGlnThrGluValHisIleuIysThrGlyValIleValPro 80
DB 411 TTGGGAGAGCAATATACATTTCCCAAGTCAAGTTCATTAAACAGGTGGATTTGTTCT 470
QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
DB 471 CCAGCAGCTCAACAGCTTCCACAGAGAAATATTCACAGAAATAGTACAGAAAGCTCTTTCT 530
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleIysProGlyLeu 120
DB 531 GCAGAGGAGACTCTCCCAAGTGAACCTCTCAGCAATTCACATCAATCAAAACGAGACTT 590
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuValLys 140
DB 591 GCTTTAAAGCTGGAGATGGCTTATATCATTTAGCTTACAGCTGGAGACAGTTAAAAAG 650
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
DB 651 CCATTCATTTCCCATTCATCATATGAGAGCTCATGCACTTACTTATAGTTGACCAATAAA 710
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyValHisCysLeuLeuAlaLeuVal 180
DB 711 GTAGATTTCCCTTTTAACTTTTGAATTTCTTGAGAGTCACTGTCGTGTTGGCAATAGTT 770
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
DB 771 CAAAGAGTTTCAGATTTTCTGCTTCTTGAAAGCTTTTGACATACATAGACACAGAGGACATG 830
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleIleHisProGluCysSerThrMetSer 220
DB 831 CTGACAGAGGTGGCAAGAGACTTTCTTTAATAAAATCAAGAGTCTCCACCATGAGT 890
QY 221 GlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
DB 891 GGTGGGAAGCCATACACATTTGGCCCAACCAAGAAATGATTTCTATTGACATATAA 950
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
DB 951 CTCCTCTTCATCATCTAATAAATTTGATTTTCTTTTACTGGAAGTTCAACAGCTTACT 1010
QY 261 AspLysIleIleIleMetLysLysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSer 280
```

```
|||||
DB 1011 GATATAAATTAATAAGAAAAGAAAAGAGAGAGTATTGAGAGGGGCAATCTCTCT 1070
QY 281 SerAlaIleAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeuValLysArg 300
DB 1071 TCAGCAGACAGCATTTGCTGCGACAGTACAGACACAAATGCAATGTCATCTGTGAAAAGA 1130
QY 301 ThrHisArgAlaIleLeuPheCysGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320
DB 1131 ACACATGGGGCTATTCTGTTTGTAGACAGAGACTTGTATACCTCAAAATATAGCATTA 1190
QY 321 LeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArgAlaLeuGluIleLeu 340
DB 1191 CTGGTTCATCTGGTGGTGTGTCGAAGTAACCTTATATCCGACAGAGCTCTGGAATTTTA 1250
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaArgLeuCysThrAspAsnGly 360
DB 1251 ACAAGCCACACACAGTCACTTGTGTGCTCTCCACACTATATGACATGATATATG 1310
QY 361 IleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
DB 1311 ATTATGATTCATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATG 1370
QY 381 IleGluGlyIleArgGlyArgProLysCysProLeuGlyValAspIleSerLysGluVal 400
DB 1371 ATGAGAGGATCCGCTATGAAACCAAAATGCTCTTGAGAGTACATATCAAAAGACTT 1430
QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
DB 1431 GGAGAGCTTCCATTAATAAGTACCAATTAATAATAATGAGATA 1472

RESULT 2
US-10-649-273-1
; Sequence 1, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OR INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649, 273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266, 518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067, 443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282, 814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
SEQ ID NO 1
LENGTH: 2197
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (231)..(1472)
US-10-649-273-1

Alignment Scores:
Pred. No.: 2,07e-247 Length: 2197
Score: 2125.00 Matches: 414
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 18 Gaps: 0

US-10-649-273-2 (1-414) x US-10-649-273-1 (1-2197)
QY 1 MetLeuIleLeuThrIleThrsThrAlaGlyValPhePheIysProSerIysArgLysValTyr 20
DB 231 ATGCTTAATCTTGACTTAAGACTGCAAGAGTTTCTTTTAAACCATCAAAAGAAAGTTTAT 290
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValLeu 40
```

DB 291 GAATTTTAAAGAGTTTAAATTTTCATCTCGAAGCACTATTCTTCAATAAATGATATG 350  
QY 41 GYIILEGLuThSerCySaPaSpThrAlaAlaValAspGluThrGlyAsnVal 60  
DB 351 GGAATTGAAACTGATTGTATGATACAGCAGCTGCTGTGTGTGATGAAATCGT 410  
QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80  
DB 411 TTGGGAGAAAGCATATCATCTCCAAAGTTCATTAAACAGGTGGATGTTCT 470  
QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
DB 471 CCAGCAGCTCAACAGCTTCACAGAAATAATTCACGAATAGTACAAAGAGCTTTCT 530  
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleArgLeuThrAsnLys 120  
DB 531 GCCAGTGGAGTCTCTCCAAAGTACCTCTCAGCAATGCACTACCATTAACCAAGACTT 590  
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140  
DB 591 GCTTTAAGCTGGAGAGTGGCTTATCATTTAGCTTACAGCTGGTGGACAGTTAAAG 650  
QY 141 ProPheIleProIleHisHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
DB 651 CCATTCAATCCCATTCATCATATGAGAGCTCATGCACTTACTATTAGTTGACCAATAAA 710  
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
DB 711 GTAGAATTTCTTTTGTAGTTGATTTCTGAGATTCGAGAGTCACTGTCTGTTGGCAATTGATT 770  
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
DB 771 CAAGAGTTTCAATTTTCTGCTTCTTGAAAGCTTTTGACATAGAACAGGTCACATG 830  
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyCysSerThrMetSer 220  
DB 831 CTTGACAGGTGGCAAGAACATTTCTTTAATAAACATCCAGAGTGTCTCCACCATGAGT 890  
QY 221 GlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
DB 891 GGGGGGAAAGCCATAGAACATTTGGCCAAACAAAGAAATGATTTCTTGTGACATCAA 950  
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
DB 951 CCGCCCTGTGATATGCTAAATAATGATTTTCTTTACTGAGCTTCAACAGCTTACT 1010  
QY 261 AspLysIleIleMetLysLysGluLysGluGlyIleGluLysGlyGlnIleLeuSer 280  
DB 1011 GATTAATTAATTAAGAAAAGAAAAGAAAGAGTATTGAGAAAGGGCAAAATCTGTCT 1070  
QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300  
DB 1071 TCGAGCAGACAGCAATGTGTCACAGTACAGACCAATGCGATGTCTCTTGAAAAAG 1130  
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGlnAsnAsnAlaVal 320  
DB 1131 ACACATCGGGCTATTCTGTTTGTGACAGAGACATTTGTAACCTCAAAATAAGCGAGTA 1190  
QY 321 LeuValAlaSerGlyValAlaLysAsnPheThrIleArgArgAlaLeuGlnIleLeu 340  
DB 1191 CTGTGTGATCTGTGTGTGTGCAAGTACTTCAATACCGCAAGCTCTGGAATTTTA 1250  
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360  
DB 1251 ACAACCGCAACACAGTCACTTGTGTGTCTCTCCAGACATATGCACTGATATGTC 1310  
QY 361 IleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
DB 1311 ATATGATTCATGAGAAATGGAATGAAGACTAGCTGTGCTGGGCAATTTTACATGAC 1370  
QY 381 IleGluGlyIleArgGlyGluProLysCysProLeuGlyValAspIleSerLysGluVal 400

DB 1371 ATAGAAGCATCCCTATGAAACCAAAATGCTCTTGTGAGTACATATCAAAAGAGTT 1430  
QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414  
DB 1431 GGAAGAGCTTCCATTAAGTACCAATTAATAATGAGATA 1472  
RESULT 3  
US-10-651-722-1  
; Sequence 1, Application US/10651722  
; Publication No. US20040048302A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
; FILE REFERENCE: D0073 DIV  
; CURRENT APPLICATION NUMBER: US/10/651,722  
; PRIOR FILING DATE: 2003-08-29  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1  
; LENGTH: 2197  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (231)..(1472)  
US-10-651-722-1  
Alignment Scores:  
Pred. No.: 2,07e-247 Length: 2197  
Score: 2125.00 Matches: 414  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0  
DB: 18  
US-10-649-273-2 (1-414) x US-10-651-722-1 (1-2197)  
QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValIle 20  
DB 231 ATGCTAATCTTGACTAGTACGACGAGTCTTTTAAACATCAAAAGAAAGCTTAT 290  
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40  
DB 291 GAATTTTAAAGATTTTAAATTTTCATCTGGAACATATTCTTCATTAATAATGATATG 350  
QY 41 GlyIleGluThrSerCySaPaSpThrAlaAlaValAspGluThrGlyAsnVal 60  
DB 351 GGAATTTGAAACTAGTTGTGATGATACAGCAGCTGTGTGTGTGATGAAACTGGAATGTG 410  
QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80  
DB 411 TTGGGAGAAAGCATATCATCTCCAAAGTTCATTAAACAGGTGGATGTTCT 470  
QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
DB 471 CCAGCAGCTCAACAGCTTCACAGAAATAATTCACGAATAGTACAAAGAGCTTTCT 530  
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleArgLeuThrAsnLys 120  
DB 531 GCCAGTGGAGTCTCTCCAAAGTACCTCTCAGCAATGCACTACCATTAACCAAGACTT 590  
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140  
DB 591 GCTTTAAGCTGGAGAGTGGCTTATCATTTAGCTTACAGCTGGTGGACAGTTAAAG 650  
QY 141 ProPheIleProIleHisHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160

Db 651 CCATTTCATCCATTCATATATGAGGCTCATGACCTTACTTATAGTTGACCAATMAA 710  
 Qy 161 ValGluPheProPheLeuValLeuLeuLeuSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
 Db 711 GAGAAATTCCTTTTATGTTAGTTCTTTGATTTCTGAGAGTACTGTCTGTTGGCAATTA 770  
 Qy 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
 Db 771 CAAGAGTTTCAGATTTTCTGCTTCTTGAGAAAGCTTTGACATAGACCAAGGTGACATG 830  
 Qy 201 LeuAspLysValAlaAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220  
 Db 831 CTGACCAAGGTGCAAGAAAGACTTTCTTATATAAAACATCAAGTCTCCACCATATAGT 890  
 Qy 221 GlyGlyLysValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
 Db 891 GGTGGAAAGCCATAGAACTTTGGCCAAACAGAAATGATTTTCAATTTGACATTA 950  
 Qy 241 ProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGlyLeuGlnHisValThr 260  
 Db 951 CTTCCCTTCATCATCTCTAAATTTGATTTTCTTTTACTGAGCTTCAACAGCTTACT 1010  
 Qy 261 AspLysIleIleMetLysLysGlyLysGlyGlyIleGluLysGlyGlnIleLeuSer 280  
 Db 1011 GATTAATAATATATGAAAAAGAAAAAGAAAGATTTGAGAAAGGGCAAAATCTGTCT 1070  
 Qy 281 SerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300  
 Db 1071 TCAGCAGCAGACATTCCTGCGCACAGTACACACACAAATGGCATGCTTTGTGAAAAA 1130  
 Qy 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320  
 Db 1131 ACACATCGGGCTATTTCTGTTTGTAGCAGAGAGACTTTACTCTCAAAATATGCACTA 1190  
 Qy 321 LeuValAlaSerGlyGlyValAlaLysAsnPheTyrIleArgArgAlaLeuGlnIleLeu 340  
 Db 1191 CTGGTTCATCTGTGTGTGTGCGCAAGTAACTTCTATATCCGAGAGCTTGGAAATTTTA 1250  
 Qy 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360  
 Db 1251 ACAAACGCAACACAGGCACTTTGTGTGCTCCCTCCGAGACTATGCACTGATTAAGGC 1310  
 Qy 361 IleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
 Db 1311 ATTATGATTCGATCGAATGTATTTGAAGACTACGTCGTGCTTGAGCATTTTATACATGAC 1370  
 Qy 381 IleGlnGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerLysGluVal 400  
 Db 1371 ATAGAAGGATCGCGCTATGAAACCAAAATGTCTCTTGAGATTAACATATCAAAAGAA 1430  
 Qy 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414  
 Db 1431 GGAGAGCTTCCATTAAGATCAACAAATATGAGATA 1472  
 RESULT 4  
 US-10-480-988-36  
 ; Sequence 36, Application US/10480988  
 ; Publication No. US20050069877A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: GANDHI, Ameena R.; KABLE, Amy E.;  
 ; APPLICANT: SWARNAKAR, Anita; HAPALIA, April J.A.;  
 ; APPLICANT: TRAN, Bao; DUGGAN, Brendan M.;  
 ; APPLICANT: WARREN, Bridget A.; ISON, Craig H.;  
 ; APPLICANT: HONCHIEL, Cynthia D.; NGUYEN, Dannel B.;  
 ; APPLICANT: LU, Dying Aina M.; LEE, Ernestine A.;  
 ; APPLICANT: YUE, Henry; FORSYTHE, Ian J.;  
 ; APPLICANT: BARBOSO, Ines; RAMKUMAR, Jayalaxmi;  
 ; APPLICANT: GRIFFIN, Jennifer A.; LI, Jena X.;  
 ; APPLICANT: YANG, Junming; THANGAVELU, Kavilcha;  
 ; APPLICANT: GIETZEN, Kimberly J.; DING, Li;  
 ; APPLICANT: BAUGHN, Mariah R.; BOROMSKY, Mark L.;  
 ; APPLICANT: YAO, Monique G.; CHAWLA, Narinder K.;  
 ; APPLICANT: MASON, Patricia M.; GURURAJAN, Rajagopal;

; APPLICANT: LEE, Sally; BECHA, Shanya D.;  
 ; APPLICANT: LEE, Soo Yeun; TRAN, Uyen K.;  
 ; APPLICANT: ELLIOTT, Vicki S.; LUD, Wen;  
 ; APPLICANT: SPRAGUE, William W.; TANG, Y. Tom;  
 ; APPLICANT: LU, Yan; ZEBARUDIAN, Yeganeh  
 ; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES  
 ; FILE REFERENCE: PF-1040 USN  
 ; CURRENT APPLICATION NUMBER: US/10/480,988  
 ; CURRENT FILING DATE: 2003-12-16  
 ; PRIOR APPLICATION NUMBER: PCT/US02/19360  
 ; PRIOR FILING DATE: 2002-06-18  
 ; PRIOR APPLICATION NUMBER: US 60/300,508  
 ; PRIOR FILING DATE: 2001-06-22  
 ; PRIOR APPLICATION NUMBER: US 60/303,445  
 ; PRIOR FILING DATE: 2001-07-06  
 ; PRIOR APPLICATION NUMBER: US 60/305,405  
 ; PRIOR FILING DATE: 2001-07-13  
 ; PRIOR APPLICATION NUMBER: US 60/311,442  
 ; PRIOR FILING DATE: 2001-08-09  
 ; PRIOR APPLICATION NUMBER: US 60/314,821  
 ; PRIOR FILING DATE: 2001-08-24  
 ; PRIOR APPLICATION NUMBER: US 60/315,992  
 ; PRIOR FILING DATE: 2001-08-29  
 ; PRIOR APPLICATION NUMBER: US 60/378,205  
 ; PRIOR FILING DATE: 2002-05-03  
 ; NUMBER OF SEQ ID NOS: 56  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 36  
 ; LENGTH: 2572  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; OTHER INFORMATION: Incyte ID No: 7632424CBI  
 US-10-480-988-36  
 Alignment Scores:  
 Pred. No.: 2,666-247 Length: 2572  
 Score: 2125.00 Matches: 414  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 21 Gaps: 0  
 US-10-649-273-2 (1-414) x US-10-480-988-36 (1-2572)  
 Qy 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20  
 Db 144 ATGCTATCTTGTACTAAGACTGACGAGAGTTTCTTTTAAACCATCAAAAGAAATTTAT 203  
 Qy 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40  
 Db 204 GAATTTTAAAGAGTTTAAATTTTTCATCTCGAACACATATTTCTTCATTAATATGATTTG 263  
 Qy 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60  
 Db 264 GGAATTTAAACTGATTTGATGATGATGACGACTCTGTGTGATGAACTGGAATGTG 323  
 Qy 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyGlyIleValPro 80  
 Db 324 TTGGGAAAGCAACATACATTTCCCAACCTGAAGTTTATTTAAACAGAGGTGGATTTGCT 383  
 Qy 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
 Db 384 CCAGACCTCAACAGCTTCCACAGAGAAATATTTCAAGAAATGATTAAGAAAGCTTTCT 443  
 Qy 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
 Db 444 GCGAGTGAAGTCTTCCAAAGTGAAGCTTCAGCAATTCATACATTAACCAAGGACTT 503  
 Qy 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140  
 Db 504 GCTTTAAGCTCGGAGTGGGCTTATTCATTTTACCTTACGCTGTGATGAGACATTAAG 563

[illegible]

|   |   |  |                      |                 |
|---|---|--|----------------------|-----------------|
|   | / | SOFTWARE:  | PatentIn version 3.0 |                 |
|   | / | SEQ ID NO 21   |                      |                 |
|   | / | LENGTH:  | 1387                 |                 |
|   | / | TYPE:  | DNA                  |                 |
|   | / | ORGANISM:  | homo sapiens         |                 |
|   | / | US-10-067-443-21   |                      |                 |
| <br>  |   |  |                      |                 |
| Alignment Scores:                                   |   |  | Pred. No.:           | 1,59e-243       |
| Score:  |   |  | 2090,50              | Length: 1387    |
| Percent Similarity:                                 |   |  | 93.85%               | Matches: 412    |
| Best Local Similarity:                              |   |  | 93.85%               | Conservative: 0 |
| Query Match:  |   |  | 98.38%               | Mismatches: 2   |
| DB:   |   |  | 14                   | Indels: 25      |
| US-10-649-273-2 (1-414) x US-10-067-443-21 (1-1387) |   |  | Gaps: 1              |                 |
| <br>  |   |  |                      |                 |
| QY  |   | 1 MetLeuIlleLeuThrlLystrHraIagIyVaI PhePheLysProSerLySarGlySValTYR | 20                   |                 |
| Db  |   | 24 ATGTAAATCTTGACTACAGACTCAGAGAGTTTTTTTAAACCATCAAAAAGGAAGTTTAT     | 83                   |                 |
| <br>  |   |  |                      |                 |
| QY  |   | 21 GluPheLeuArgSerPheasnPhelSRProGIYThLeuPheLeuhIsIstIleVallLeu    | 40                   |                 |
| Db  |   | 84 GAATTTTAAAGAAGTTTTTAATTTTCATCCCGAACAACATACTTCTCATATAAAATGATTG   | 143                  |                 |
| <br>  |   |  |                      |                 |
| QY  |   | 41 GIYIIegIuThSeCySaSpasPrThrIaaIaIaValaValaSPGLuThrGIyASnVal      | 60                   |                 |
| Db  |   | 144 GGAAITGGAAACTAGTTGTGCATGATACAGACGCTGCTGTGTGATGAATGAATGTG       | 203                  |                 |
| <br>  |   |  |                      |                 |
| QY  |   | 61 IeuGIyGuIaAIleHisSerGIlnThrgIuValIHISLeuLYSThrGIyIIeValPro      | 80                   |                 |
| Db  |   | 204 TTGGAGGAAGCAATACATCCCACAACTGAAGTTCAITTTAAAAACAGGTGGATTGTTCCT   | 263                  |                 |
| <br>  |   |  |                      |                 |
| QY  |   | 81 ProAlaIaIaIngIlnLeuHISarPGIuaenIIegIlnArGIleValGIngIuaIaleuSer  | 100                  |                 |
| Db  |   | 264 CCAGAGCGTCMACACGCTTCACAGAGAAATATTCACGAATATGATACAGAGCTCTTTCT    | 323                  |                 |
| <br>  |   |  |                      |                 |
| QY  |   | 101 AlaSerGIyValSerProSerAspLeuSerAlaIleaIaThrTrILeUYSProGIyleu    | 120                  |                 |
| Db  |   | 324 GCCAGTGGAGTCTCTCCAAAGTACCTCTTCAGCATTTGCCAATCACATAAACAGAGACTT   | 383                  |                 |
| <br>  |   |  |                      |                 |
| QY  |   | 121 AlaleuSerLeuGIyValGilyLeuSerPheSerLeuGIlnleuValGilylnleuLYS    | 140                  |                 |
| Db  |   | 384 GCTTTAAGCGCTGGAGTGGGCTTATCATTTAGCTTCACCTGGTAGACAGTTAAAAAG      | 443                  |                 |
| <br>  |   |  |                      |                 |
| QY  |   | 141 PropHeIIeProIIeHIshISmetGIuaIHISaIaLeuThrlIeaIroLeuThrAsnLYS   | 160                  |                 |
| Db  |   | 444 CCATTCATTCCTCATTCATCATATGAGGCTGCATGACCTTACATTAAGTGTGACCAATAAA  | 503                  |                 |
| <br>  |   |  |                      |                 |
| QY  |   | 161 ValGIuPheProPheLeuValIleuLeuIIeSerGIyVIHISCYSyLeuLeuAlaleuVal  | 180                  |                 |
| Db  |   | 504 GTAGATTTCTTTTTTATTGTTCTTTTGATTTCGAGAGTCACGTCTGTGGCATTAAGTT     | 553                  |                 |
| <br>  |   |  |                      |                 |
| QY  |   | 181 GIngIyValSerAspPheLeuLeuLeuGIyLYSLeuAspIIleaIAProGIyaSPMet     | 200                  |                 |
| Db  |   | 564 CAAGAGTTTCAGATTTTCTGCTCTTCGAAAAGTCTTGTGACATACACACAGGTGCATG     | 623                  |                 |
| <br>  |   |  |                      |                 |
| QY  |   | 201 IeuAspLySValAlaIargIroLeuSerIleuIHISaIaProGIuCYSeSrTHImetSer   | 220                  |                 |
| Db  |   | 624 CTGTGACMAGTGTGCAGAGACTCTTTTAATAAAMCATCCAGAGTCTCACACATGAGT      | 683                  |                 |
| <br>  |   |  |                      |                 |
| QY  |   | 221 GIYGIyLYSAIaIleGIuHISLeuAlalySGINGIyASArGPheHISpHeasPIIelys    | 240                  |                 |
| Db  |   | 684 GGTGGAAAGCCATAGAGCATTTGGCCMAACAAGAAATAGATTTCATTTTGACATCAA      | 743                  |                 |
| <br>  |   |  |                      |                 |
| QY  |   | 241 ProProLeuHISHISaIalalySAnCYsaAPheSerPheThrGIyleuGInHISvalThr   | 260                  |                 |
| Db  |   | 744 CCGCCCTTCATCANGCTAAAAATGTGATTTTCTTTACTGAGACTTCAACAGTTACT       | 803                  |                 |
| <br>  |   |  |                      |                 |
| QY  |   | 261 AspySIIeIIemELySLySGIuLYSGIuGIuGIyIIe-----                     | 273                  |                 |
| Db  |   | 804 GATTAATAATATATAGAAAAAGAAAAAGGAAGGATATATTCTTAATAGTAAAGTTGAA     | 863                  |                 |

QY 274 -----Glulys 275  
DB 864 CAGATAAATATTCCTGATTCGCTTAAAAATAGCTGCTCATTTCTGCAGGTATGAGAG 923  
QY 276 GtGlnlleuSerSerAlaIAspIleAlaIaThrValGlnHsThrMetAlaCys 295  
DB 924 GGGCAAAATCTGCTTCAGACGACGACATTCGTCACAGACGACACAAATGGCATTGT 983  
QY 296 HisLeuValIysArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuPro 315  
DB 984 CATCTGTGAAAGAACACATCGGGCTATCTGTTTGTAAAGCAGAGACTTCTTACT 1043  
QY 316 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaAsenPheTyrlleArgArg 335  
DB 1044 CAAATAATATGACATGCTGTTGCATCTGGTGTCGCAAGTAACCTCTATATCCGAGA 1103  
QY 336 AlaleuGlnlleuThrAsnAlaThrGlnCysThrleuLeuCysProProProArgLeu 355  
DB 1104 GCTCTGAAATTTTAAACAAACGACACAGTGCATTTGTGTCTCTCCCAAGACTA 1163  
QY 356 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 375  
DB 1164 TGCATGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1223  
QY 376 GlyIleLeuHisAspIleGluGlyIleArgTyrlleGluProIysCysProLeuGlyValAsp 395  
DB 1224 GGCATTTTACATGACATGAGAGCATCGCTATGAAACAAATGCTCTCTTGAGTAC 1283  
QY 396 IleSerIleGluValGlyGluAlaSerIleLeuValProGlnleuIysMetGluIle 414  
DB 1284 ATATCAAAAGAAAGTTGAGAGAGCTTCATTAAGTACACAAATTAAGAGATA 1340

## RESULT 6

US-10-649-273-21

Sequence 21, Application US/10649273  
Publication No. US2004043407A1

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1

FILE REFERENCE: D0073 CMT

CURRENT APPLICATION NUMBER: US/10/649, 273

CURRENT FILING DATE: 2003-08-27

PRIOR APPLICATION NUMBER: US 60/266, 518

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 10/067, 443

PRIOR FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/282, 814

NUMBER OF SEQ ID NOS: 71

SOFTWARE: PatentIn version 3.2

SEQ ID NO 21

LENGTH: 1387

TYPE: DNA

ORGANISM: homo sapiens

US-10-649-273-21

## Alignment Scores:

| Pred. No.:             | 1,596-243 | Length:       | 1387 |
|------------------------|-----------|---------------|------|
| Score:                 | 2090.50   | Matches:      | 412  |
| Percent Similarity:    | 93.85%    | Conservative: | 0    |
| Best Local Similarity: | 93.85%    | Mismatches:   | 2    |
| Query Match:           | 98.38%    | Indels:       | 25   |
| DB:                    | 18        | Gaps:         | 1    |

US-10-649-273-2 (1-414) x US-10-649-273-21 (1-1387)

QY 1 MetLeuIleLeuThrIysThrAlaGlyValPhePheIysProSerIysArgIysValTyr 20  
DB 24 ATGCTAATCTTGCTAAAGACTGACGAGGTTTTTTTTTAAACCATCAAAAGAAAGTTTAT 83  
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrIleuPheLeuHisIleValIleu 40  
DB 84 GAATTTTAAAGAGTTTATTTTATTTTCAATCCGAAACACTATTCTTCATTAATAATGATATG 143

QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThrGlyAsnVal 60  
DB 144 GAAATGAAACTAGTGTGATGATGATACAGAGCTCTGTGGATGAAATGGAATATGTG 203  
QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisIleuIysThrGlyIleValPro 80  
DB 204 TTGGGAAGACAAATACATTCCTCAAACTGAAGTTTATTAATAAGAGGGGATGTTCT 263  
QY 81 ProAlaIleGlnGlnIleuHisArgGluAsnIleGlnArgIleValGlnAlaLeuSer 100  
DB 264 CACGACCTCAACAGCTTCACAGAGAAATTTTCAAGATATGACAAAGAGCTTTCT 323  
QY 101 AlaserGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleIysProGlyLeu 120  
DB 324 GCGAGTGAAGTCTCCCAAGTACCTTCAGCATTTCACTAACATTAACCAAGACTT 383  
QY 121 AlaleuSerleuGlyValGlyLeuSerPheSerleuGlnleuValGlyGlnleuIys 140  
DB 384 GCTTAAAGCCCTGGAGTGGCTTATCATTTAGCTTACAGCTGTAGGACAGTTAAAG 443  
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnIys 160  
DB 444 CCATTCAATCCCATTCATCATATGAGAGCTTCATGACTTACTATTAGGTTGACCAATAA 503  
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
DB 504 GTAGAAATTCCTTTTATTTAGTTCTTTGATTTCTGAGAGTCACTGTCTGTGGATTAGT 563  
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyIysSerleuAspIleAlaProGlyAspMet 200  
DB 564 CAAGAGATTTCAGATTTCTGCTCTTGGAAAGCTTGGACATAGCACAGTGAACATG 623  
QY 201 LeuAspIysValAlaIaArgArgLeuSerleuIleIysHisProGluCysSerThrMetSer 220  
DB 624 CTTGACAAAGTGGGAAAGAGCTTTCTTTAATAAATCAACAGTGTCCACCATGAGT 683  
QY 221 GlyGlyIysAsnAlaIleGluHisIleuAlaIysGlnGlyAsnArgPheHisPheAspIleIys 240  
DB 684 GGTGGAAAGCCATAGAGCATTTTGGCCAAACAGAAATAGATTTCATTTTGACATCAAA 743  
QY 241 ProProLeuHisHisAlaIysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
DB 744 CTTCCCTTGATCATGCTAAATAATTTGATTTCTTTTACTGACCTTCAACAGTTACT 803  
QY 261 AspIysIleIleMetIysGlyGluGlyIle-----GluIys 273  
DB 804 GATTAATAATTAATGAAGAAAGAAAGAGAGATATTTCTTAATTAGTAAAGTTGAA 863  
QY 274 -----GluIys 275  
DB 864 CAGATAAATATTCCTGATTCGCTTAAAAATAGCTGCTCATTTCTGCAGGTATGAGAG 923  
QY 276 GtGlnlleuSerSerAlaIAspIleAlaIaThrValGlnHsThrMetAlaCys 295  
DB 924 GGGCAAAATCTGCTTCAGACGACGACATTCGTCACAGACGACACAAATGGCATTGT 983  
QY 296 HisLeuValIysArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuPro 315  
DB 984 CATCTGTGAAAGAACACATCGGGCTATCTGTTTGTAAAGCAGAGACTTGTACT 1043  
QY 316 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaAsenPheTyrlleArgArg 335  
DB 1044 CAAATAATATGACATGCTGTTGCATCTGGTGTCGCAAGTAACCTCTATATCCGAGA 1103  
QY 336 AlaleuGlnlleuThrAsnAlaThrGlnCysThrleuLeuCysProProProArgLeu 355  
DB 1104 GCTCTGAAATTTTAAACAAACGACACAGTGCATTTGTGTCTCTCCCAAGACTA 1163  
QY 356 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 375  
DB 1164 TGCATGATATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1223

QY 376 GYILELEUHIASAPILLEGUGLYILEARGTYGIDUPROLYSCYSPROLEUGLYVALASP 395  
DB 1224 GGCAATTTTACATGATGAGAGGATCGGCTATGATGACCAAAATGCTCTTGAGATGAC 1283  
QY 396 ILESERLYSGUVALGUGLYUASERILLYSVALPROGIDLEULYMEUGLUILE 414  
DB 1284 ATATCAAAAGAGGTGGAGAGGCTTCATTAAGTACCAATTAATAATGAGATA 1340

RESULT 7  
US-10-651-722-21  
; Sequence 21, Application US/10651722  
; Publication No. US20040048302A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1  
; FILE REFERENCE: D0073 DIV  
; CURRENT APPLICATION NUMBER: US/10/651,722  
; PRIOR FILING DATE: 2003-08-29  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 21  
; LENGTH: 1387  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-651-722-21

Alignment Scores:  
Pred. No.: 1,596-243 Length: 1387  
Score: 2090.50 Matches: 412  
Percent Similarity: 93.85% Conservative: 0  
Best Local Similarity: 93.85% Mismatches: 2  
Query Match: 98.38% Indels: 25  
DB: 18 Gaps: 1

US-10-649-273-2 (1-414) x US-10-651-722-21 (1-1387)

QY 1 MetleuileuThrlyrThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20  
DB 24 ATGCTAATCTTGACTAAGACTGACGAGAGTCTTTTAAACCATCAAAAGAAAGTTTAT 83  
QY 21 GlupheLysSerPheAsnPheHisProGlyThrLeuPheLysIleValLeu 40  
DB 84 GAATTTTAAAGAGTTTAAATTTTCATCTGAAACACTAATTTCTTCAATTAATGATATG 143  
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAspGluThrGlyAsnVal 60  
DB 144 GGAATTAAGAACTATTGATGATGATACAGCAGCTGCTGTGGTGAAGAACTGAAATG 203  
QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80  
DB 204 TTGGAGAGAGCAATACATTTCCCAAACTGAAGTTCTTAAACACGGTGGAGATGTTCT 263  
QY 81 ProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
DB 264 CCGACGACCTCAACAGCTTCACAGAGAAATATTCACAGAAATGATACAGAGAGCTTTCT 323  
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
DB 324 GCGAGTGAAGTCTTCCAGAGGAGCTTCACAGAAATGACATACATTAACCAAGAGATT 383  
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140  
DB 384 GCTTAAGCTGGAGAGTGGCTTATCATTTAGCTTACAGCTGTGAGACAGTAAAG 443  
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
DB 444 CCATTCATTCCTCATTCATCATATGAGAGGCTCATGACATCTACTATATGAGTGAACAT 503

QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
DB 504 GTAGAAATTTCTTTTATGTTCTTTGATTTCTGGAGGTCACTGCTGTGGCATTAATG 563  
QY 181 GlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
DB 564 CAAGAGATTTGAGATTTCTCTCTCTTGGAAAGCTTTTGGATATGACACAGGTACATG 623  
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyCysSerThrMetSer 220  
DB 624 CTGACCAAGGTGGCAGAGAGCTTTCTTTAATAAATCCAGAGTCTCCACCATGAGT 683  
QY 221 GlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
DB 684 GGTGGAGAGCATGAGACATTTGGCCAAACAGAGAAATAGTTTCATTTGACATCAAA 743  
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
DB 744 CTTCCCTTGATCATGCTAATAAATTTGATTTTCTTTTACTGACATTCACACCTTACT 803  
QY 261 AspLysIleIleMetLysLysGluLysGluGlyIle-----GluLys 273  
DB 804 GATAAATTAATTAATAAAGAAAGAAAGAGATATTTCTTAATTAAGTAAGTTGAA 863  
QY 274 -----GluLys 275  
DB 864 CAGATTAATATTTCTCGATGCTGCTAATAATAGCTGCTATTCTGCAGAGTATGAGAG 923  
QY 276 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 295  
DB 924 GGGCAATCTCTGTTTACAGACAGACAGATGCTGCCACAGTACACACATGAGCATGT 983  
QY 296 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 315  
DB 984 CATCTTGAAAGAAACATCGGCTATTTCTGTTTGAACAGAGACATGTTTACT 1043  
QY 316 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArg 335  
DB 1044 CAAATTAATGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1103  
QY 336 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 355  
DB 1104 GCTCTGAAATTTTAAACCAACCAACAGATGACATTTGTTGTTGTTGTTGTTGTTG 1163  
QY 356 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 375  
DB 1164 TGCACTGATTAATGCAATATGATTCATGAGAAATGATTAAGAAAGCTACAGTGTGGCTTG 1223  
QY 376 GlyIleLeuHisAspIleGluGlyIleArgTyrgIuProLysCysProLeuGlyValAsp 395  
DB 1224 GGCAATTTTAAATGATGATGAGAGGATCCGCTATGAAACCAAAATGCTCTTTGAGATGAC 1283  
QY 396 ILESERLYSGUVALGUGLYUASERILLYSVALPROGIDLEULYMEUGLUILE 414  
DB 1284 ATATCAAAAGAGGTGGAGAGGCTTCATTAAGTACCAATTAATAATGAGATA 1340

RESULT 8  
US-10-887-553A-1047  
; Sequence 1047, Application US/10887553A  
; Publication No. US20050085436A1  
; GENERAL INFORMATION:  
; APPLICANT: Garza, Dan  
; APPLICANT: Li, Hao  
; TITLE OF INVENTION: Method to treat conditions associated  
; TITLE OF INVENTION: with insulin signalling dysregulation  
; FILE REFERENCE: 4-33262  
; CURRENT APPLICATION NUMBER: US/10/887,553A  
; PRIOR FILING DATE: 2004-07-08  
; PRIOR APPLICATION NUMBER: 60/485,883  
; PRIOR FILING DATE: 2003-08-07  
; NUMBER OF SEQ ID NOS: 1208  
; SOFTWARE: FastSeq for Windows Version 4.0

| SEQ ID NO 1047   | LENGTH: 1387 | TYPE: DNA   | ORGANISM: human | US-10-887-553A-1047 |
|--|--------------|---|-----------------|---------------------|
| Alignment Scores:                                      |              |   |                 |                     |
| Pred. No.:   | 1.59e-243    | Length:   | 1387            |                     |
| Score:   | 2090.50      | Matches:  | 412             |                     |
| Percent Similarity:                                    | 93.85%       | Conservative:   | 0               |                     |
| Best Local Similarity:                                 | 93.85%       | Mismatches:   | 2               |                     |
| Query Match:   | 98.38%       | Indels:   | 25              |                     |
| DB:  | 21           | Gaps:   | 1               |                     |
| US-10-649-273-2 (1-414) x US-10-887-553A-1047 (1-1387) |              |   |                 |                     |
| QY   | 1            | MetLeuIleuThrIlySthralaGlyValPhePheIysProSerIlySarGlyValIlyr      | 20              |                     |
| DB   | 24           | ATGTAATCTTGACTAGACAGCTGAGAGATTTTAAAAACATCAAAAGGAAGTTAT            | 83              |                     |
| QY   | 21           | GIuPheIeuIaYserPheAsnPhenHisProGlyThrIeuPheIeuHisIlySileValIeu    | 40              |                     |
| DB   | 84           | GAATTTTAAAGATTAAATTTTCACTCCGAAACACTATTCTTCAATAAATAGTATG           | 143             |                     |
| QY   | 41           | GIlyIlegIuThrIserCyAspAspThrIalalalalalValIaIaSPGIuThrGIyAsnVal   | 60              |                     |
| DB   | 144          | GGAATTTGAACCTAGTGTGATGATGATACAGACGCTCTGTGGTATGAAATCGAAATGTG       | 203             |                     |
| QY   | 61           | LeuGIyGluIalalIeHisSerGIuThrGIuValHisIleuIySthrGIyGIlyIleValPro   | 80              |                     |
| DB   | 204          | TTGGAGAGAGCAATACATCTCCCAACTGAAGCTCTTTAAAAACAGGTGGATTTGTTCT        | 263             |                     |
| QY   | 81           | ProIalalagInIleuHisIarGIuIasnIlegInarGIleValGIuIuIalIeuSer        | 100             |                     |
| DB   | 264          | CCACAGAGCTCAACAGCTTCACAGAGAAATATATCAAGAAATGATCAAGAAAGCTTTCT       | 323             |                     |
| QY   | 101          | AlaSerGIyValSerProSerAspIeuSerIalIlealIthrThrIleIyProGIyIleu      | 120             |                     |
| DB   | 324          | GCCAGTGAAGTCTCTCCAAAGTGAACCTTCACCAATTCGACTCCATTAACCAAGAGCTT       | 383             |                     |
| QY   | 121          | AlaIeuSerIeuGIyValGIyIeuSerPheSerIeuGIuIeuValGIyGIuIeuIySlys      | 148             |                     |
| DB   | 384          | GCTTTAAGCCCTGGAGAGTGGCTTATCATTTAGCTTACAGCTGTGAGACATTTAAAAAG       | 443             |                     |
| QY   | 141          | ProPheIleProIleHisIleMetGIuIalHisIalIeuThrIlearGIeuThrAsnIlys     | 160             |                     |
| DB   | 444          | CCATTCAATCCCATTCATCATATGAGAGGCTCATGACTACTATTAAGTTGACCAATAAA       | 503             |                     |
| QY   | 161          | ValGIuPheProPheIeuValIleuIleIseerGIyGIyHisIySyleuIeuIalIeuVal     | 180             |                     |
| DB   | 504          | GTAAGATTTCTTTTAAATTTAGTTCTTTGATTTCTGAGAGTCACTGTCTGTGGCAATTGTT     | 563             |                     |
| QY   | 181          | GIuGIyValIserAspPheIeuIeuIeuGIyIySerIeuAspIleIalProGIyAspMet      | 200             |                     |
| DB   | 564          | CAAGAGATTCAAGATTTTCTCTCTTGGAAAGCTTTTGGACATAGCACAGGTGAATG          | 622             |                     |
| QY   | 201          | IeuAspIySValIalIarIarIeuSerIeuIleIyHisIProGIyCySerThrMetSer       | 220             |                     |
| DB   | 624          | CTTGCAAGAGTGGCAAGAGACTTTCTTTTAAATAAACATCCAGAGTGTCTCCACATGAGT      | 683             |                     |
| QY   | 221          | GIyGIyIySValIalIegIuHisIeuIalIyGIuGIuIySnaIyPheHisIspHeAspIleIyS  | 240             |                     |
| DB   | 684          | GGTGGGAAAGCCATAGACATTTTGGCCAAACAAAGAAATATGATTTCTTTTGGACATCAA      | 743             |                     |
| QY   | 241          | ProProIeuHisIalIalIyAsnCyAspPheSerPheThrGIyIeuGIuHisIalIyThr      | 260             |                     |
| DB   | 744          | CTCTCCCTTGCAATAGCTTAATAATTTGATTTTCTTTTACTGAGACTTCAACAGCTTACT      | 803             |                     |
| QY   | 261          | AspIySileIleMetIyIyGIuIyGIuIyIleIleIleIleIleIleIleIleIleIleIleIle | 273             |                     |
| DB   | 804          | GATTAATAATATATGAAAGAAAGAAAGAAAGAGAGATATATTTCTTAATTAGTAAGTTGAA     | 863             |                     |
| QY   | 274          | -----GIuIyS 275   |                 |                     |

|    |      |   |      |
|----|------|---|------|
| Dd | 864  | CAGATAAATATTCTCCGATGTGGCCTAAATAATAGTCGCTCATTTCTGCAGGATATGAAGAAG | 923  |
| Oy | 276  | GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys    | 295  |
| Dd | 924  | GCGCAATATCTGTCTTTCAGACAGACAATTCGTCACAGTACAGCACACAATATGCATGT     | 983  |
| Oy | 236  | HisteValIleValArgThrHiseArgAlaIleLeuPheCysGlyGlnArgAspLeuLeuPro | 315  |
| Dd | 984  | CATCTTGTAAGAAAAGAACACATCGGGCTATTCTGTTTTGTAAACAGAGACACTTGTTAACT  | 1043 |
| Oy | 316  | GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrlleArgArg    | 335  |
| Dd | 1044 | CAAAATATATGCACTACGTGTGCATCTGTGTGTGCAGAGTAACCTTCTATATCCGACAG     | 1103 |
| Oy | 336  | AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu       | 355  |
| Dd | 1104 | GCTCGGAATTTTAAACAAAGCAAACAAGTGCACATTTGTGTGTGTCCTCCACAGACTA      | 1163 |
| Oy | 356  | CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu    | 375  |
| Dd | 1164 | TGCACGTATATATGGCATATGATATGTCATCGAATGTGATTTGAAGAAGCTAGCTGTGCTTGG | 1223 |
| Oy | 376  | GlyIleLeuHisAspIleGluGlyIleArgArgArgLysProLysCysProLeuGlyValAsp | 395  |
| Dd | 1224 | GGCATTATTACATGATATGAAGAGCATCGCTATGTAAACCAAAATGTCTCTTGGAGTAGAC   | 1283 |
| Oy | 396  | IleSerLysGlyValGlyAlaSerIleLysValProGlnLeuLysMetGluIle          | 414  |
| Dd | 1284 | ATATCAAAGAGAGTGGAGAAAGCTTCCATTAAGATGCCACATTTAAATAATGGAGATAT     | 1340 |

```

RESULT 9
US-10-012-140-6
; Sequence 6, Application US/100122140
; Publication No. US20030009017A1
; GENERAL INFORMATION:
; APPLICANT: Leiby, Kevin R.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Glucksmann, Maria A.
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USSES
; TITLE OF INVENTION: THERSOP
; FILE REFERENCE: 381552004900
; CURRENT APPLICATION NUMBER: US/10/012,140
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/246,768
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,772
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,185
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1245
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-012-140-6

Alignment Scores:
Pred. No.:          2.7e-243          Length:          1245
Score:              2088.00           Matches:          407
Percent Similarity: 99.03%           Conservative:     3
Best Local Similarity: 98.31%         Mismatches:      4
Query Match:        98.26%           Indels:           0
DB:                 14               Gaps:             0

US-10-649-273-2 (1-414) x US-10-012-140-6 (1-1245)

Cy          1 MetcunllleauThlyeThAlaGlyValPhePhelyProSerlySarGlyValTyr 20
Db          1 ATGCTAACTCTGACCTAAGACTGACGAGAGTCTTTTAAACCATCAAAAGGAAAGCTTTAT 60

```



QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValLeu 40  
 DB 61 GAATTTTAAGAAGTTTAAATTTTCACTCTGGAACACATATTTCTTCAATAAATAGTATG 120  
 QY 41 GlyTlIleGluThrSerCysAspAspThrAlaAlaAlaValAspGluThrGlyAsnVal 60  
 DB 121 GGAATTGAAACTGATGTGTGATGATACAGCAGCTGCTGTGTGTGTAAGAACTGGAAATGTG 180  
 QY 61 LeuGluGluAlaIleHisSerGlnThrGluValHisIleuLeuThrGlyIleValPro 80  
 DB 181 TTGGAGAAAGCAATACATTCCTCCAACTGAAGTTTCAATTTAAACAGGTGGATTTGTTCT 240  
 QY 81 ProAlaAlaGlnGlnLeuHisIleArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
 DB 241 CCAGAGCTCAACAGCTTCCAGAGAAATATTCAGAAATAGTACAAAGAGCTTTTCT 300  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleuLeuProGlyLeu 120  
 DB 301 GCCAGTGAAGTCTCTCCAACTGACCTCTCAGCAATTCACCACTAAACAGAGACTT 360  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLeuLeu 140  
 DB 361 GCTTTAAGCCTGGAGAGGCTTATATCTTACCTTACAGCTGTAGAGCAGTTAAAG 420  
 QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLeu 160  
 DB 421 CCATTCATTCCTCATTCATATGAGAGCTCATGCACTTACATTAAGTTACCAATAA 480  
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
 DB 481 GTAGAAATTCCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 540  
 QY 181 GlnGlyValSerAspPheLeuLeuGlyLeuSerLeuAspIleAlaProGlyAspMet 200  
 DB 541 CAAGAGTTTCAGATTTTCTGCTTCTTGAAGCTTTTGAAGCTTTTGAAGCTTTTGAAGCT 600  
 QY 201 LeuAspLeuValAlaArgArgLeuSerLeuIleHisIleProGluCysSerThrMetSer 220  
 DB 601 CTTGACAAAGGTGGCAAGAGACTTCTTCTTAAATAAATCCAGAGTCTCCACATGAGT 660  
 QY 221 GlyGlyAsnAlaIleGluHisIleuAlaLeuGlnGlyAsnArgPheHisPheAspIleLeu 240  
 DB 661 GTGGGAAAGCCATAGAACATTTTGGCCAAACAGAAATATATTTTCAATTTTCAATCAATCAA 720  
 QY 241 ProProLeuHisIleHisAlaLeuAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
 DB 721 CTTCCCTTGATCATGCTAAATAATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 780  
 QY 261 AspLeuValIleMetLeuIleGluGluGluGluGluGluGluGluGluGluGluGluGlu 280  
 DB 781 GATTAATAATATATGAAACAGGAAACAGGAAAGGATATTTGAAAGGGGCAATCTGTCT 840  
 QY 281 SerAlaAlaAspIleAlaIleAlaThrValGlnHisIleThrMetAlaCysHisIleValLeuArg 300  
 DB 841 TCGACAGACGATTTGCTGCTGACAGTACAGCAACATAGCATGTCATCTTGTAAAGAA 900  
 QY 301 ThrHisArgAlaIleLeuPheCysLeuGlnArgAspLeuLeuProGlnAsnAlaVal 320  
 DB 901 ACACATCGGCTATTTCTTTTGTAGAGAGAGACTTTTACCTCAAAATATATGCACTA 960  
 QY 321 LeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArgAlaLeuGluIleLeu 340  
 DB 961 CTGGTTCATCTGGTGGTGGCAAGTAACTTATATATCCGACAGAGCTCTGCAAAATTTTA 1020  
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360  
 DB 1021 ACAAACCAACACAGTGCATTTGTGTGTCTCTCCACAGCTATATGCACTGATATATG 1080  
 QY 361 IleMetIleAlaTPAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsn 380  
 DB 1081 ATTATGATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 445  
 QY 381 IleGluGlyIleArgGlyGluProLeuCysAspLeuGlyValAspIleSerLeuGluVal 400

DB 1141 ATGAAAGCATCCCTATGAAACAAATGTCTCTTGAGATGACATATCAAAAGAGTT 1200  
 QY 401 GlyGluAlaSerIleLeuValProGlnLeuLeuMetGluIle 414  
 DB 1201 GGAGAGCTTCCATTAATAAGTACCAATTAATAATGAGATA 1242  
 RESULT 10  
 US-10-012-140-4  
 ; Sequence 4, Application US/10012140  
 ; Publication No. US2003009017A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Leidy, Kevin R.  
 ; APPLICANT: Kapeller-Libermann, Rosana  
 ; APPLICANT: Glucksmann, Maria A.  
 ; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND  
 ; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES  
 ; TITLE OF INVENTION: THEREOF  
 ; FILE REFERENCE: 38155204900  
 ; CURRENT APPLICATION NUMBER: US/10/012,140  
 ; PRIOR FILING DATE: 2001-11-08  
 ; PRIOR APPLICATION NUMBER: 60/246,768  
 ; PRIOR FILING DATE: 2000-11-08  
 ; PRIOR APPLICATION NUMBER: 60/246,772  
 ; PRIOR FILING DATE: 2000-11-08  
 ; PRIOR APPLICATION NUMBER: 60/249,185  
 ; PRIOR FILING DATE: 2000-11-15  
 ; NUMBER OF SEQ ID NOS: 49  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 1820  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (146)...(1390)  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)...(1820)  
 ; OTHER INFORMATION: n = A,T,C or G  
 ; US-10-012-140-4  
 Alignment Scores:  
 Pred. No.: 4,91e-243 Length: 1820  
 Score: 2088.00 Matches: 407  
 Percent Similarity: 99.03% Conservative: 3  
 Best Local Similarity: 98.31% Mismatches: 4  
 Query Match: 98.26% Indels: 0  
 DB: 14 Gaps: 0  
 US-10-649-273-2 (1-414) x US-10-012-140-4 (1-1820)  
 QY 1 MetLeuIleLeuThrIleThrAlaGlyValPhePheLeuProSerIleArgIleValIle 20  
 DB 146 ATGCTAATCTTGACATCAAGCTGCAAGAGTTTCTTAAACATCAAAAAAGAAAGTTAT 205  
 QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValLeu 40  
 DB 206 GAATTTTAAGAAGTTTAAATTTTCACTCTGGAACACATATTTCTTCAATAAATAGTATG 265  
 QY 41 GlyTlIleGluThrSerCysAspAspThrAlaAlaAlaValAspGluThrGlyAsnVal 60  
 DB 266 GGAATTTGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 325  
 QY 61 LeuGluGluAlaIleHisSerGlnThrGluValHisIleuLeuThrGlyIleValPro 80  
 DB 326 TTGGAGAAAGCAATACATTCCTCCAACTGAAGTTCAATTTAAACAGGTGGATTTTCT 385  
 QY 81 ProAlaAlaGlnGlnLeuHisIleArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
 DB 386 CCAGAGCTCAACAGCTTCCAGAGAAATATTCAGAAATAGTACAAAGAGCTTTTCT 445  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleuLeuProGlyLeu 120



```
Db 446 GCCAGTGGAGTCTCTCCAGTGAACCTCTCAGCAATTCACATTAACCAACGAGATT 505
Qy 121 AAlaSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140
Db 506 GCTTTAAGCTGGGAGGAGGCTTATCATTTAGCTTACAGCTGGTAGGACAGTTAAAAAG 565
Qy 141 ProPheIleProIleHisIleMetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
Db 566 CCATTCAATCCCATTCATCATATGAGAGGCTCATCTACTATTAAGTTGACCAATPAA 625
Qy 161 ValGlnPheProPheLeuValIleLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
Db 626 GATGAATTCCTTTTATGATTTTATTTATTTGAGAGTCACTGCTGTTGGCATTAGTT 685
Qy 181 GlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
Db 686 CAAAGAGTTTCAATTTTCTGCTCTTGGAAAGCTTTGACATACACACAGTGAATG 745
Qy 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysIleProGlyCysSerThrMetSer 220
Db 746 CTGGACAAAGGTGCAAGAGACTTTCTTTATATAAATCATCCAGAGTGTCTCCACATGAGT 805
Qy 221 GlyGlyLysAlaIleGlyHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
Db 806 GGTGGGAAACCCATGAAACATTTGGCCAAACAGAAATGATTTTCATTTTGACATCAA 865
Qy 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
Db 866 CTTCCCTTGATCATAGCTAAATTTGATTTTCTTTTACTGAGCTTCAACAGTTACT 925
Qy 261 AspLysIleIleMetLysGlyLysGlyGlnGlyIleGlyLysGlyGlnIleLeuSer 280
Db 926 GATTAATAATATGATAAAGAGAGAAACAGAGAGAGATTTGAGAGAGGGAATCTGCT 985
Qy 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300
Db 986 TCAGACGACAGACATTTGCTGCACAGTACAGACACAAATGTCATCTTTGTGAAAAA 1045
Qy 301 ThrHisArgAlaIleLeuPheCysLysGlnArgSerLeuProGlnAsnAlaVal 320
Db 1046 AACATCGGGGCTATCTGTTTGTGAGCAGAGACATTTGTTACTTCAAAATAAGCGAGTA 1105
Qy 321 LeuValAlaSerGlyValAlaSerAspPheThrIleArgArgAlaLeuGlnIleLeu 340
Db 1106 CTGGTTGCATCTGCTGTGTCGCAAGTAACTTCTATATCCGCAAGCTCTGGAATTTTA 1165
Qy 341 ThrAsnAlaThrGlnCysThrLeuLeuCysAspProProArgLeuCysThrAspAsnGly 360
Db 1166 ACAAAACGCAACACAGTGCATTTGTTGTGTCCTCCACAGACTATGCACTGATTAATGCG 1225
Qy 361 IleMetIleAlaThrPangGlyIleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
Db 1226 ATTATATATTCATGAGATGATGATTTGAAAGACTAGTCTGCTGGCATTTTTCATGAC 1285
Qy 381 IleGlnGlyIleArgGlyArgProLysCysProLeuGlyValAspIleSerLysGlyVal 400
Db 1286 ATGGAAGGCTCCGCTATGAAACCAAAATGCTCTTTGAGTAGCAATTCAAAAGAGCTT 1345
Qy 401 GlyGlnAlaSerIleLysValProGlnLeuLysMetGlnIle 414
Db 1346 GGAAGAGCTTCCATTAAGTAAACCAATTAAGTAAATGAGAGTA 1387

RESULT 11
US-10-094-749-400
; Sequence 400, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
```

```
APPLICANT: ISHII, SHIZUKO
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUKIO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHICO
APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTOMYUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
FILE OF INVENTION: NOVEL FULL-LENGTH CDNA
FILE REFERENCE: 084335/0160
CURRENT APPLICATION NUMBER: US/10/094,749
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350,435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328381
PRIOR FILING DATE: 2001-09-14
NUMBER OF SEQ ID NOS: 3381
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 400
LENGTH: 2208
TYPE: DNA
ORGANISM: Homo sapiens
US-10-094-749-400

Alignment Scores:
Pred. No.: 2,26e-225 Length: 2208
Score: 1944.00 Matches: 386
Percent Similarity: 93.24% Conservative: 0
Best Local Similarity: 93.24% Mismatches: 4
Query Match: 91.48% Indels: 24
DB: 17 Gaps: 1

US-10-649-273-2 (1-414) x US-10-094-749-400 (1-2208)
Qy 1 MetLeuIleLeuThrIleValGlyValPhePheLysProSerLysArgLysValTyr 20
Db 344 ATGCAATTCCTGACTTAAGACATGACAGAGCTTTTAAACCAATCAAAAGAAAGTTAT 403
Qy 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40
Db 404 GAATTTTAAAGATTATTTATTTTCACTCTGGAACATATTTCTTCAATAAATAGATTG 463
Qy 41 GlyIleGlnThrSerCysAspAspThrAlaAlaAlaValAlaAspGluThrGlyAsnVal 60
Db 464 GGAATTGAATCTAGTTGATGATGATACACAGCTGCTGTGTGATGAAGAACTGGAATGTG 523
Qy 61 LeuGlyGlnAlaIleHisIleSerGlnThrGlyValHisLeuLysThrGlyGlyIleValPro 80
Db 524 TTGGAGAGAGCAATACATTTCCCAACTGAAGTTCATTTAAAAACAGTGGAGATTGCT 583
Qy 81 ProAlaAlaGlnGlnLeuHisArgGlyAsnIleGlnArgIleValGlnGlnAlaLeuSer 100
Db 584 CAGCAGCTTCAACACCTTCAACAGAGAAATATTTCAACAAATGATGACAGAGCTCTTTGT 643
Qy 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
Db 644 GCCAGTGAAGTCTCTCCAAAGTACCTCTCAGAAATGGAATGCAATCAATAAACAGAGACTT 703
Qy 121 AAlaSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140
Db 704 GCTTTAAGCTGGGAGGCTTATCATTTAGCTTACAGCTGGTAGGACAGTTAAAAAG 763
Qy 141 ProPheIleProIleHisIleMetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
Db 764 CCATTCAATCCCATTCATCATATGAGAGGCTCATCTACTATTAAGTTGACCAATPAA 823
Qy 161 ValGlnPheProPheLeuValIleLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
```

```

Db      824 GTAGAAATTCCTTTTGTAGTTCTTTGATTTCTGGAGCTACTCTGTGGCAATTAGTT 883
Qy      181 GInGIyValSerAspPheLeuLeuLeuGIyLysSerLeuAspIleAlaProGIyAspMet 200
Db      884 CAAAGAGTTTCAGATTTTCTGCTTCTTGAAAAGCTTTGGACATAGCACAGGAGCATG 943
Qy      201 LeuAspLysValAlaArgArgLeuSerLeuIleLysIleProGIyCysSerThrMetSer 220
Db      944 CTGGACAAAGGTGGCAAGAAAGACTTCTTATATAAATCATCCAGAGTCTCCACCATGAGT 1003
Qy      221 GIyGIyLysValAlaIleGIyLysLeuAlaLysGIyGIyLysAspPheIlePheAspIleLys 240
Db      1004 GGTGGAAAAGCCATAGAAACATTTGGCCAAACAAAGAAATATATTTCAATTTGACATCAAA 1063
Qy      241 ProProLeuHisIleAlaLysAsnCysAspPheSerPheThrGIyLeuGIyLysValThr 260
Db      1064 CTTCCCTTGATCATGCTAATAAATTTGATTTTCTTTACTTGACCTTCAACAGCTTACT 1123
Qy      261 AspLysIleIleMetLysLysGIyLysGIyLysGIyLysGIyLysGIyLysGIyLysGIyLys 280
Db      1124 GATTAATATATATATGAAAAGAAAAGAAAAGAGATTTGAAAGGGGCAATCTGTCT 1183
Qy      281 SerAlaAlaAspIleAlaAlaIleValAlaIleIleThrMetAlaCysHisLeuValLysArg 300
Db      1184 TCAGCAGACAGACATTCCTGCCACAGTACAGACACAAATGCGATGATCTTTGTGAAAAGA 1243
Qy      301 ThrHisArgAlaIleLeuPheCysLysGIyLysArgAspLeuLeuProGIyAsnAspAlaVal 320
Db      1244 ACAATCTGGGCTATTCGTTTGTGTAACAGAGACATTTGTAACCTCAAAATATGAGTA 1303
Qy      321 LeuValAlaSerGIyGIyValAlaSerAspPheThrIleArgArgAlaLeuGIyLysLeu 340
Db      1304 CTGTTGTCATCTGGTGTGTGTCAGAGTAACTTCTGTATCCGACAGCTCTGGAATTTTA 1363
Qy      341 ThrAsnAlaThrGIyCysThrLeuLeuCysProProArgLysCysThrAspAsnGIy 360
Db      1364 ACAAGCCAAACAGTCACTTGTGTGTCTCTCCACAGCATATGACATGATATGCG 1423
Qy      361 IleMetIleAlaTrpAsnGIyLysLeuArgLeuArgAlaGIyLeuGIyLysLeuHisAsp 380
Db      1424 ATATGATTCGA-----TGATGCTCTCTGGAGTAAACATATCAAAAGAGTT 1435
Qy      381 IleGIyLysIleArgTrpGIyLysProLysCysProLeuGIyValAspIleSerLysLeuVal 400
Db      1436 -----TGATGCTCTCTGGAGTAAACATATCAAAAGAGTT 1471
Qy      401 GIyGIyLysSerIleLysValProGIyLysMetGIyLys 414
Db      1472 GGAGAGCTTCCATTAAGTAACCAATTAATAAATGAGATTA 1513

RESULT 12
US-10-723-860-7447
; Sequence 7447, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnick, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723, 860
; PRIOR FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7447
; LENGTH: 2890
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature

```

```

; LOCATION: (646)..(657)
; OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-7447

Alignment Scores:
Pred. No.: 3,45E-225 Length: 2890
Score: 1944.00 Matches: 386
Percent Similarity: 93.24% Conservative: 0
Best Local Similarity: 93.24% Mismatches: 4
Query Match: 91.48% Indels: 24
DB: 20 Gaps: 1

US-10-649-273-2 (1-414) x US-10-723-860-7447 (1-2890)
Qy      1 MetLeuIleLeuThrIleThrAlaGIyValPhePheLysProSerLysArgLysValThr 20
Db      1001 ATGCTAATCTTGACTAAGACCTGACAGAGTTTCTTTTAAACATCAAAAAGAAATTTAT 1060
Qy      21 GluPheLeuArgSerPheAsnPheHisProGIyThrLeuPheLeuHisLysIleValLeu 40
Db      1061 GAATTTTAAAGAGTTTAAATTTTCACTCGAACAATAATTTCTTCAATTAATAATAGTATG 1120
Qy      41 GIyLysIleGIyThrSerCysAspAspThrAlaAlaValAlaAspGIyThrGIyLysVal 60
Db      1121 GGAATTTAAAGTATGTTGATGATGATACAGACACTGCTGTGTGATGAAACTGGAAATGTG 1180
Qy      61 LeuGIyLysAlaIleHisSerGIyThrGIyLysIleLysLeuValHisLeuLysThrGIyLysVal 80
Db      1181 TTGGAGAAAGCAATACATCTCCAACTGAAGTTCACTTTAAACACAGGTGGATTTCTCT 1240
Qy      81 ProAlaAlaGIyLysLeuHisArgGIyLysAlaIleGIyLysValGIyLysLeuValLeuSer 100
Db      1241 CCAGAGCTCAACAGCTTCAACAGAAATTTCAACGAATGTACAAAGCTCTTTCT 1300
Qy      101 AlaSerGIyValSerProSerAspLeuSerAlaIleAlaThrIleLysProGIyLysLeu 120
Db      1301 GCCAGTGAAGTCTCTCCAAAGTGAACCTTCAGCAATTTGCAACTACATTAACCAAGACTT 1360
Qy      121 AlaLeuSerLeuGIyValGIyLysSerPheSerLeuGIyLysValGIyLysLeuLysLys 140
Db      1361 GCTTTAAAGCTGGAGGTGGCTTATCATTTAGCTTACAGCTGTGTGAGAGATTAATAAAG 1420
Qy      141 ProPheIleProIleHisIleMetGIyLysAlaIleAlaLeuThrIleArgLeuThrAsnLys 160
Db      1421 CCATTCATTTCCCATTCATCATATGAGGCTCATGACATTAATAGTTGACCAATAAA 1480
Qy      161 ValGluPheProPheLeuValLeuLeuIleSerGIyLysHisCysLeuLeuAlaLeuVal 180
Db      1481 GTAGAAATTCCTTTTGTAGTTCTTTGATTTCTGGAGGTCACTGTCTGTGGCATTAGTT 1540
Qy      181 GInGIyValSerAspPheLeuLeuLeuGIyLysSerLeuAspIleAlaProGIyAspMet 200
Db      1541 CAAGAGATTTCAAGATTTCTGCTTCTTGAAGTCTTTGAGATGACACAGGTACATG 1600
Qy      201 LeuAspLysValAlaArgArgLeuSerLeuIleLysIleProGIyCysSerThrMetSer 220
Db      1601 CTGCAAGAGTGGCAAGAAAGCTTCTTATATAAACATCCAGAGTCTCCACCATGACT 1660
Qy      221 GIyGIyLysValAlaIleGIyLysLeuAlaLysGIyLysAspPheHisPheAspIleLys 240
Db      1661 GGTGGAAAAGCCATAGAAACATTTGGCCAAACAGAAATATGATTTCAATTTTCAATCAAA 1720
Qy      241 ProProLeuHisIleAlaLysAsnCysAspPheSerPheThrGIyLeuGIyLysValThr 260
Db      1721 CTTCCCTTGATCATGCTAATAAATTTGATTTTCTTTTACTGGAATTTCAACAGCTTACT 1780
Qy      261 AspLysIleIleMetLysLysGIyLysGIyLysGIyLysGIyLysGIyLysGIyLysGIyLys 280
Db      1781 GATTAATAATATATATGAAAAGAAAAGAAAAGAGATTTGAGAAAGGATTAATCTGTCT 1840
Qy      281 SerAlaAlaAspIleAlaAlaIleValAlaIleIleThrMetAlaCysHisLeuValLysArg 300
Db      1841 TCAGCAGCAGACATTCCTGCCACAGTACAGACACAAATGCGATGATCTTTGTGAAAAGA 1900

```

QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGlnAsnAsnAlaVal 320  
Db 1901 ACACATGGGCTATTCTGTTTGTAGCAGAGACCTTGTAACCAATAATATGCACTA 1960  
QY 321 LeuValAlaSerGlyValAlaSerAsnPheTyrIleArgAlaIleGluIleLeu 340  
Db 1961 CTGGTGGCATCTGGTGGTGGTGGCAAGTACCTTCTGTATCCGCAAGCTTGGAAATTTTA 2020  
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuGlyThrAspAsnGly 360  
Db 2021 ACMAAGCCACACAGTGCACCTTGTGTGTCTCTCCAGACTATGCACTGATATATGCG 2080  
QY 361 IleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
Db 2081 ATTATGATTGCA----- 2092  
QY 381 IleGluGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerLysGluVal 400  
Db 2093 -----TGATGTCTCTTGGAGTAGACATATCAAAAGAGTT 2128  
QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414  
Db 2129 GGAGAGCTTCCATMAAAGTACCAATMAAATGAGATA 2170  
RESULT 13  
US-10-120-988-177  
; Sequence 177, Application US/10120988  
; Publication No. US20030219745A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Goodrich, Ryje  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Ren, Feiyun  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Dimañac, Radoje T.  
; TITLE OF INVENTION: No. US20030219745A1el Nucleic Acids and  
; FILE OF INVENTION: Polypeptides  
; FILE REFERENCE: 802CON  
; CURRENT APPLICATION NUMBER: US/10/120,988  
; CURRENT FILING DATE: 2002-04-11  
; PRIOR APPLICATION NUMBER: 09/774,528  
; PRIOR FILING DATE: 2001-01-30  
; NUMBER OF SEQ ID NOS: 441  
; SOFTWARE: Pf\_Fl\_genes Version 2.0  
; SEQ ID NO 177  
; LENGTH: 1416  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (205)..(1305)  
US-10-120-988-177  
Alignment Scores:  
Pred. No.: 1,07e-201 Length: 1416  
Score: 1747.00 Matches: 340  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.71% Mismatches: 0  
Query Match: 82.21% Indels: 0  
DB: 17 Gaps: 0  
US-10-649-273-2 (1-414) x US-10-120-988-177 (1-1416)  
QY 74 LysThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArg 93  
Db 280 AGAACAGGTGGAGTTGTTCTCCAGCAGCTCAACAGCTTCAAGAAATATTCAACGA 339  
QY 94 IleValGlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAla 113  
Db 340 ATAGTACAAAGAGCTTCTTCTGCGAGTGGAGTCTCTCAATGACCTCTCAGCAATTTCA 399  
QY 114 ThrThrIleLysProGlyLeuAlaLeuSerIleuGlyValGlyLeuSerPheSerLeuGln 133

Db 400 ACTACCAATAAACCAGGACTTGCTTAAGCCCTGGAGAGGGCTTATCATTTAGCTTACG 459  
QY 134 LeuValGlyGlnLeuLysLysProPheIleProIleHisMetGluAlaHisAlaLeu 153  
Db 460 CTGGTAGACAGCTTAAAGAACCATTTCTTCCATTCATATATGAGGCTCATGCATCT 519  
QY 154 ThrIleArgLeuThrAsnLysValGluPheProPheLeuValLeuLeuIleSerGlyGly 173  
Db 520 ACTATTAGTGGACCAATTAAGTAGAATTTCCCTTTTATGTTCTTTGATTTCTGGAGGT 579  
QY 174 HisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeu 193  
Db 580 CACTGTCTGTTGGCATTTGTTCAAGAGTTTCAGATTTTCCTTCTTGGAAAGCTTTTG 639  
QY 194 AspIleAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHis 213  
Db 640 GACATAGCACCAAGGAGCATGCTTGACAAAGGTGGCAAGAGACTTTCTTAATAAACAAT 699  
QY 214 ProGluCysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsn 233  
Db 700 CCAAGTCTCTCCACATGAGTGGTGGGAAGCCATAGAACATTTGGCCAAACAAGGAAT 759  
QY 234 ArgPheHisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPhe 253  
Db 760 AGATTTCATTTTGACATCAAACTCCCTTGATCATGCTAAAAATTTGATTTTCTTTT 819  
QY 254 ThrGlyLeuGlnHisValThrAspLysIleIleMetLysLysGlnGluGlyIle 273  
Db 820 ACTGACCTTCAACAGGTTACTGATTAATTAATGAAGAAAGAGAAAGAGAGATTT 879  
QY 274 GluLysGlyGlnIleLeuSerSerAlaIleAspIleAlaIleThrValGlnHisIleThrMet 293  
Db 880 GAGAAAGGCGCAATCTCTGTTCTTCCAGCAGACATTTGCTCCACAGTACAGACCAAG 939  
QY 294 AlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeu 313  
Db 940 GCATGTCATCTTGTGAAGAAACACATCGGGCTATTCTGTTTGTAGCAGAGACTTG 999  
QY 314 LeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIle 333  
Db 1000 TTACTCTCAAAATTAAGCAGTACTGTGTCTATCTGTGTGTGTGTGTGTGTGTGTGT 1059  
QY 334 ArgArgAlaLeuGlnIleLeuThrAsnAlaIleThrGlnCysThrLeuLeuCysProPro 353  
Db 1060 CGCAGAGCTCTGGAAATTTTAACCAACGACACAGTGCATTTGTGTCTCTCTCC 1119  
QY 354 ArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAla 373  
Db 1120 AGACTATGCACATGATTAAGGCAATTATGATGATGGAATGGTAAAGACTACGTGCT 1179  
QY 374 GlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGly 393  
Db 1180 GGCCTTGGCATTTTACATGACATAGACAGCAATCCCTATGAACCAAAATGTCTCTTGA 1239  
QY 394 ValAspIleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGlu 413  
Db 1240 GTAGCATATCAAAAGAAAGTTGAGAGGCTTCCATTAAGTACCACAATTAATAATGAG 1299  
QY 414 Ile 414  
Db 1300 ATA 1302  
RESULT 14  
US-10-067-443-23  
; Sequence 23, Application US/10067443  
; Publication No. US20030082782A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443

```

; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
; US-10-067-443-23

Alignment Scores:
Pred. No.: 1,41e-157 Length: 1526
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 65,18% Indels: 0
DB: 14 Gaps: 0

US-10-649-273-2 (1-414) x US-10-067-443-23 (1-1526)
QY 148 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 167
DB 1 ATGAGGCTCATGCACTTACTATTAGTTGACCAATAAGTAGAATTTCTTTTAGTT 60
QY 168 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 187
DB 61 CTTTGATTTCTGAGGCTCATGCTCTGTCATTAGTTCAAGAGATTTCCAGATTTCTG 120
QY 188 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 207
DB 121 CTTCTGGAAGCTCTTGGACATGACACAGGTGACATGCTTGACAAAGTGGCAAGGA 180
QY 208 LeuSerLeuIleLysHisProGlyLysCysSerThrMetSerGlyGlyLysAlaIleGlnHis 227
DB 181 CTTTCTTTAATAAACAATCCAGAGTCTCCACCAATGATGTTGGGAAGCCATGAAACAT 240
QY 228 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProPheLeuHisAlaLys 247
DB 241 TTGGCCAAACAAAGGAATAGATTTCAATTTGACATCAACCTCCCTTGACATGCTTAA 300
QY 248 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 267
DB 301 AATTGATTTCTTGGAGGCTCATGCTTCAACACGTTCTGATTAATATATGAAAG 360
QY 268 GluLysGlnGlyLysIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 287
DB 361 GAAAAAGAGAGAGTATTGAGAGGGGCAATCTGCTTCAGACAGACATGCTGCC 420
QY 288 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 307
DB 421 ACGTACAGACACAATGGCATGTCATCTTGTGAAGAACACATCGGCTATTTCTGTT 480
QY 308 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 327
DB 481 TGTAAAGAGAGAGCTTGTAACTCAAAATATATGACATGCTGTCATGCTGCTGTC 540
QY 328 AlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaIleMetLysLys 347
DB 541 GCAAGTAACCTTATATATCCGAGAGCTCTGGAATTTTAAACAAAGCAACAGTGCACT 600
QY 348 LeuLeuCysProPheProArgLeuCysThrAspAsnGlyIleMetIleAlaTriPasnGly 367
DB 601 TTGTTGTCCTCTCCCAAGCATATGACATATATGCAATTAATGATTCATGCAATGCT 660
QY 368 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGlu 387
DB 661 ATTGAAGAAGCATAGTGTGCTGGGCAATTTACATGACATAGAAGGATCGGCTATGA 720
QY 388 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 407
DB 407 |||||
```

```

DB 721 CCAAAATGCTCTTGGAGTAGACATATCAAAAGAGTTGGAGAGCTTCCATAAAGTA 780
QY 408 ProGlnLeuLysMetGluIle 414
DB 781 CCACATTTAAAAATGAGAGATA 801

RESULT 15
US-10-649-273-23
; Sequence 23, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEINASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
; US-10-649-273-23

Alignment Scores:
Pred. No.: 1,41e-157 Length: 1526
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 65,18% Indels: 0
DB: 18 Gaps: 0

US-10-649-273-2 (1-414) x US-10-649-273-23 (1-1526)
QY 148 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 167
DB 1 ATGAGGCTCATGCACTTACTATTAGTTGACCAATAAGTAGAATTTCTTTTAGTT 60
QY 168 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 187
DB 61 CTTTGATTTCTGAGGCTCATGCTCTGTCATTAGTTCAAGAGATTTCCAGATTTCTG 120
QY 188 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 207
DB 121 CTTCTGGAAGCTCTTGGACATGACACAGGTGACATGCTTGACAAAGTGGCAAGGA 180
QY 208 LeuSerLeuIleLysHisProGlyLysCysSerThrMetSerGlyGlyLysAlaIleGlnHis 227
DB 181 CTTTCTTTAATAAACAATCCAGAGTCTCCACCAATGATGTTGGGAAGCCATGAAACAT 240
QY 228 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProPheLeuHisAlaLys 247
DB 241 TTGGCCAAACAAAGGAATAGATTTCAATTTGACATCAACCTCCCTTGACATGCTTAA 300
QY 248 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 267
DB 301 AATTGATTTCTTGGAGGCTCATGCTTCAACACGTTCTGATTAATATATGAAAG 360
QY 268 GluLysGlnGlyLysIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 287
DB 361 GAAAAAGAGAGAGTATTGAGAGGGGCAATCTGCTCTTACAGACAGACATGCTGCC 420
QY 288 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 307
DB 421 ACGTACAGACACAATGGCATGTCATCTTGTGAAGAACACATCGGCTATTTCTGTT 480
QY 308 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 327
DB 327 |||||
```

```
Db      ||||| 481 TGTAAACGAGAGACTGTGTTACCTCAAAATATGCAAGTACTGGTGCATCTGGTGTC 540
Qy      ||||| 328 AlaSerAsnPhenylleuArgAlaleuGluilleuThrAsnAlaThrGlnCysThr 347
Db      ||||| 541 GCAGATTAACCTTCTATATCCGACAGACTCTGGAAATTTTACAAACGACACAGTGCACT 600
Qy      ||||| 348 LeuLeuCysProProProArgLeuCysThrAspAsnGlylleMetIleAlaTyrPasnGly 367
Db      ||||| 601 TTGTTGTGTCTCTCCACAGACTATGCACTGATATGCAATTGATTCATGCAATGAT 660
Qy      ||||| 368 IleGluArgLeuArgAlaGlyLeuGlylleuHisAspIleGluGlylleuArgTyrGlu 387
Db      ||||| 661 ATTGAAGAAGCTACGTGCTGGCTTGGCATTTTACATGACATAGAGGCAATCCGCTATGAA 720
Qy      ||||| 388 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 407
Db      ||||| 721 CCAAAATGTCCTTGGAGTAGACATATCAAAAGAGGTGGAGAGCTTCATTAAGAATA 780
Qy      ||||| 408 ProGlnLeuLysMetGluIle 414
Db      ||||| 781 CCACATTAATAAATGAGATA 801
```

Search completed: June 17, 2005, 08:03:08  
Job time : 2842.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: June 16, 2005, 15:19:45 ; Search time 3193.26 Seconds  
(Without alignments)  
4051.513 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_148\_414  
Perfect score: 1385  
Sequence: 1 MEAHATIRLTKVKEPPFLV.....DISKEVGEASIKVPQLMEI 267

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues  
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=xlp  
-Q/cgna2\_1/USPTO.spool\_p/US10649273/runat\_15062005\_111416\_6030/app\_query.fasta\_1.1429  
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0 -LIST=45  
-INITs=bits -START=1 -END=-1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pco -NOR=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10649273 @CGN 1.1 7742 @runat\_15062005\_111416\_6030 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:  
1: gb ba:\*  
2: gb htg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_dr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID       | Description        |
|------------|-------|-------------|--------|----------|--------------------|
| 1          | 1385  | 100.0       | 1416   | AR541929 | AR541929 Sequence  |
| 2          | 1385  | 100.0       | 1526   | AR428809 | AR428809 Sequence  |
| 3          | 1385  | 100.0       | 1908   | BC011904 | BC011904 Homo sapi |
| 4          | 1385  | 100.0       | 2197   | AR428803 | AR428803 Sequence  |

|    |        |      |        |    |             |                    |
|----|--------|------|--------|----|-------------|--------------------|
| 5  | 1358.5 | 98.1 | 1387   | 6  | AR428808    | AR428808 Sequence  |
| 6  | 1358.5 | 98.1 | 1387   | 9  | HS295148    | AU295148 Homo sapi |
| 7  | 1348   | 97.3 | 1245   | 6  | AX664697    | AX664697 Sequence  |
| 8  | 1348   | 97.3 | 1820   | 6  | AX664695    | AX664695 Sequence  |
| 9  | 1231   | 88.9 | 1844   | 10 | BC058172    | BC058172 Mus muscu |
| 10 | 1224   | 88.4 | 1017   | 10 | BC038910    | BC038910 Mus muscu |
| 11 | 1204   | 86.9 | 2208   | 6  | AX713716    | AX713716 Sequence  |
| 12 | 1204   | 86.9 | 2208   | 9  | AK055441    | AK055441 Homo sapi |
| 13 | 1116   | 80.6 | 1546   | 10 | BC078974    | BC078974 Rattus no |
| 14 | 950.5  | 68.6 | 1109   | 10 | BC051211    | BC051211 Mus muscu |
| 15 | 939    | 67.8 | 1522   | 5  | BX934991    | BX934991 Gallus ga |
| 16 | 938    | 67.7 | 860    | 5  | BX930694    | BX930694 Gallus ga |
| 17 | 934    | 67.4 | 1558   | 5  | BX930693    | BX930693 Gallus ga |
| 18 | 831.5  | 59.0 | 84115  | 9  | AC013468    | AC013468 Homo sapi |
| 19 | 824.5  | 59.5 | 14365  | 6  | AR428807    | AR428807 Sequence  |
| 20 | 666    | 48.1 | 249601 | 2  | AC114153    | AC114153 Rattus no |
| 21 | 666    | 48.1 | 308652 | 2  | AC121478    | AC121478 Rattus no |
| 22 | 652.5  | 47.1 | 256751 | 10 | AC122925    | AC122925 Mus muscu |
| 23 | 610    | 44.0 | 875    | 6  | CQ721898    | CQ721898 Sequence  |
| 24 | 488.5  | 35.3 | 1474   | 3  | AK113378    | AK113378 Ciona int |
| 25 | 477.5  | 34.5 | 121251 | 5  | AL591593    | AL591593 Zebrafish |
| 26 | 435    | 31.4 | 117322 | 5  | AL672217    | AL672217 Zebrafish |
| 27 | 415.5  | 30.0 | 1576   | 3  | AY051882    | AY051882 Drosophi  |
| 28 | 415.5  | 30.0 | 1601   | 6  | CQ606432    | CQ606432 Sequence  |
| 29 | 415.5  | 30.0 | 3656   | 6  | CQ606431    | CQ606431 Sequence  |
| 30 | 415.5  | 30.0 | 14679  | 2  | AC018262    | AC018262 Drosophi  |
| 31 | 415.5  | 30.0 | 180263 | 3  | AC010671    | AC010671 Drosophi  |
| 32 | 415.5  | 30.0 | 207432 | 3  | AE003513    | AE003513 Drosophi  |
| 33 | 408    | 29.5 | 1672   | 8  | AK070912    | AK070912 Oryza sat |
| 34 | 407.5  | 29.4 | 1443   | 8  | AY024338    | AY024338 Arabidops |
| 35 | 407.5  | 29.4 | 1474   | 8  | AY117283    | AY117283 Arabidops |
| 36 | 407.5  | 29.4 | 1567   | 8  | AY063864    | AY063864 Arabidops |
| 37 | 401.5  | 28.0 | 1557   | 2  | AY084577    | AY084577 Arabidops |
| 38 | 389    | 28.1 | 110000 | 2  | AP006501_08 | Continuation (9 of |
| 39 | 369    | 26.6 | 260271 | 1  | AE017258    | AE017258 Wolbachia |
| 40 | 356.5  | 25.7 | 333800 | 1  | SMES91792   | AL591792 Sinorhizo |
| 41 | 356    | 25.7 | 20021  | 10 | AF367967S1  | AF367967 Mus muscu |
| 42 | 356    | 25.7 | 179252 | 10 | AF131205    | AF131205 Mus muscu |
| 43 | 352    | 25.4 | 349116 | 1  | AP003003    | AP003003 Mesorhizo |
| 44 | 349    | 25.2 | 110000 | 1  | AE011797_01 | Continuation (2 of |
| 45 | 342    | 24.7 | 756    | 6  | AX685201    | AX685201 Sequence  |

#### ALIGNMENTS

RESULT 1  
LOCUS AR541929 1416 bp DNA  
DEFINITION Sequence 177 from patent US 6743619.  
ACCESSION AR541929  
VERSION AR541929.1 GI:53934009  
KEYWORDS  
SOURCE  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1416)  
Tang,Y.T., Zhou,P., Goodrich,R., Liu,C., Aeundi,V., Ren,F.,  
AUTHORS Zhang,J., Zhao,Q.A., Yang,Y., Xue,A.J., Wehtman,T., Wang,J.-R.,  
Wang,D. and Drmanac,R.T.  
TITLE Nucleic acids and polypeptides  
JOURNAL Patent: US 6743619-A 177 01-JUN-2004;  
FEATURES  
source Location/Qualifiers  
1..1416  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Alignment Scores:  
Pred. No.: 1.37e-126  
Score: 1385.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 100.00%  
Matches: 1416  
Mismatches: 267  
Indels: 0

| DB:  | 5  | Gaps:   | 0      |
|--|--|---|--------|
| US-10-649-273-2_COPY_148_414 (1-267) x AR541929 (1-1416) |  |   |        |
| QY   | 1  | MetGluA1aHisAlaLeuThr11LeaArgLeuThrAsnLysVal1GluPheProPheLeuVal   | 20     |
| Db   | 502  | ATGAGAGGCTCATGCACTTACTATTAGTTAGTTGACCAATAAAGTAAAGATTTCCTTTTAAAGTT | 561    |
| QY   | 21   | LeuLeu11LeSerG1yG1yHisCysLeuLeuA1aLeuVal1G1ng1yValSerAspPheLeu    | 40     |
| Db   | 562  | CTTTTGAATTTTCGAGAGGTCACGTCTGTGGTGGCATTAAGTCAAGAGAGTTTCAGATTTCGTG  | 621    |
| QY   | 41   | LeuLeuG1yLysSerLeuAsp11a1a1aProG1yAspMetLeuAspLysVal1a1aArgArg    | 60     |
| Db   | 622  | CTTTCTTGAAAGTCTTTGGACATGACCAAGTGCATGCTTGAACAAGTGCAGAAAGA          | 681    |
| QY   | 61   | LeuSerLeu11LeYHis1aProG1yCysSerThrMetSerG1yG1yLysAla11eG1yHis     | 80     |
| Db   | 682  | CTTTCTTTAAATAAACAATCCAGAGTGCCTCCACATGAGTGGTGGGAAAGCATAGAACAT      | 741    |
| QY   | 81   | LeuA1aLysG1ng1yAsnArgPheHisPheAsp11eLysProProLeuHis1a1aLys        | 100    |
| Db   | 742  | TTGGCCCAAAACAAGAGAAATAGATTTTCATTTTACATCAAACTCCCTTGATCAATGCTAAA    | 801    |
| QY   | 101  | AsnCysAspPheSerPheThrG1yLeuG1nHis1a1aThrAspLys11e1aMetLysLys      | 120    |
| Db   | 802  | AAATGTGATTTTCTTTTACTGTGACTTCAACACGTTACTGTAAATAATATATGAAAAAG       | 861    |
| QY   | 121  | G1uLysG1ng1yG1y11eG1uLysG1yG1n11eLeuSerSer1a1aAsp11a1a1a          | 140    |
| Db   | 862  | GAAAAAAGAGAGAGATTTGAGAAAGGGCAAAATCTGTCTTCAAGACAGACATTCCTGCC       | 921    |
| QY   | 141  | ThrVal1G1nHisThrMetA1aCysHisLeuVal1aLysArgThrHisArgAla11eLeuPhe   | 160    |
| Db   | 922  | ACAGTACAGCACACAATGACATGTCATCTTGTGAAAAAGAACACATCGGGCTATTCGTGTT     | 981    |
| QY   | 161  | CysLysG1nArgAspLeuLeuProG1nAsnAsnA1aValLeuVal1a1aSerG1yG1yVal     | 180    |
| Db   | 982  | TGTAAAGCAGAGAGACTGTGTAACTCCAAATATATACAGTACGTGTGACTGTGGTGTCTC      | 1041   |
| QY   | 181  | AlaSerAsnPheYr11eArgArgAlaLeuG1n11eLeuThrAsnA1aThrG1nCysThr       | 200    |
| Db   | 1042   | GCAAGTAACTTCTATATCCGAGAGCTTGGAAATTTTAAACAAACGACACAGTGCACT         | 1101   |
| QY   | 201  | LeuLeuCysProProProArgLeuCysThrAspAsnG1y11eMet11eAla1a1a1a1a1a1a   | 220    |
| Db   | 1102   | TTGTGTGTCTCTCTCCAGACTATGACCTGATATGACATTAATGATGATGATGATGATGAT      | 1161   |
| QY   | 221  | 11eG1uArgLeuA1aG1yLeuG1y11eLeuHisAsp11eG1uG1y11aArgTyrGlu         | 240    |
| Db   | 1162   | ATTGAAAAAGACTACACGTCTGTGGCATTGTTTAAATGACATAGAAAGGATCCGCTATGAA     | 1221   |
| QY   | 241  | ProLysCysProLeuG1yVal1aAsp11eSerLysG1uVal1aG1yG1uA1aSer11eLysVal  | 260    |
| Db   | 1222   | CCAAATAGTCTCTTGGAGTACATATCAAAAGAAAGTGGAGAAAGCTTCATTAAGTA          | 1281   |
| QY   | 261  | ProG1nLeuLysMetG1u11e267  |        |
| Db   | 1282   | CCACAAATTAATAAATGAGAGATA 1302                                     |        |
| RESULT 2   |  |   |        |
| LOCUS  | AR428809   | 1526 bp   | DNA    |
| ACCESSION  | Sequence 23  | from patent US 6642041.   | linear |
| VERSION  | AR428809   |   |        |
| KEYWORDS   | AR428809.1   | GI:40188595   |        |
| SOURCE   | Unknown.   |   |        |
| ORGANISM   | Unknown.   |   |        |
| REFERENCE  | Unclassified.  |   |        |
| AUTHORS  | Chen,J., Feder,J.N., Nelson,T.C., Kytecek,S.R. and Duclos,F. |   |        |
| TITLE  | Polynucleotides encoding a novel metalloprotease, Mp-1       |   |        |

|  |  |
|--|--|
| JOURNAL  | Patent : US 6642041-A 23 04-Nov-2003;                                  |
| FEATURES   | Location/Qualifiers  |
| Source   | 1..1526  |
|  | /organism="unknown"  |
| ORIGIN   | /mol_type="genomic DNA"  |
| Alignment Scores:  |  |
| Pred. No.:   | 1.5e-126   |
| Score:   | 1385.00  |
| Percent Similarity:                                      | 100.00%  |
| Best Local Similarity:                                   | 100.00%  |
| Query Match:   | 100.00%  |
| DB:  | 6  |
|  | Gaps: 0  |
| US-10-649-273-2_COPY_148_414 (1-267) x AR428809 (1-1526) |  |
| Qy   | 1 MetGuaAaHisAlaLeuThrIleArgLeuThrAsnLysValAlaGluPheProPheLeuVal 20    |
| Db   | 1 ATGAGAGGCTCATGACCTACTACTATTGATGGTTGACCAATAAGTAAGTAATTCCTTTTAGTT 60   |
| Qy   | 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40     |
| Db   | 61 CTTTGTGATTTCTGGAGGTCACCTGTCGTGGTGGCATTAATTCAGAGGAGTTTCAGATTTCG 120  |
| Qy   | 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgAs 60      |
| Db   | 121 CTTCTTGGAAAGCTCTTTGGACATACGACCAAGGTGACATCTTGGACAGGTGGCAAGAA 180    |
| Qy   | 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyLysAlaIleGluHis 80     |
| Db   | 181 CTTTCTTTAATATAAACATCCAGAGTGCTCCACCATGAGTGGTGGGAAAGCCATAGAACAT 240  |
| Qy   | 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100    |
| Db   | 241 TTGGCCCAACCAAGGAATAAGATTATTTGACATCAACCCCTTGATCATATGCTTAA 300       |
| Qy   | 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120   |
| Db   | 301 AATTGTGATTTTCTTTTACTGACCTTCAACGCTTACTGATTAATAATTAATGAAAAAG 360     |
| Qy   | 121 GluLysGlnGlnGlyIleGluLysGlyGlnIleLeuSerSerAlaIleAspIleAlaIle 140   |
| Db   | 361 GAAAAAGAGGAAGGATTTGAGAAAGGGCAAAATCCTGTCTTCAGCAGCAGACATTCGCCC 420   |
| Qy   | 141 ThrValGlnHisThrMetAlaCysHisIleuValLysArgThrHisArgAlaIleLeuPhe 160  |
| Db   | 421 ACAGTACGCGACACACATGGCAATGTATCTTGGAAAGAAACATCGGGCTAATTCGTTT 480     |
| Qy   | 161 CysLysGlnAAspAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGlyGlyVal 180 |
| Db   | 481 TGTAAAGCAGAGAGACTTGTACTTCAAAATAATGCAAGTACTGTTGCACTGTGGTGTGC 540    |
| Qy   | 181 AlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200   |
| Db   | 541 GCAAGTACTCTTATATCCGCGAGAGCTCTGGAAATTTTAAACAAACGCAACACAGTGCAC 600   |
| Qy   | 201 LeuLeuCysProProProArgGlnCysThrAspAsnGlyIleMetIleAlaTyrAsnGly 220   |
| Db   | 601 TTGTTGTGTCTCTCCCAAGACTTAAGCACTGATTAATGAGCATTAATGATTCAGGAATGG 660   |
| Qy   | 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGlu 240   |
| Db   | 661 ATTGAAAGACATACGTCGTGGCTTGGCATTTTACATGATGATGAAGAGCATCCGCTATGAA 720  |
| Qy   | 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260   |
| Db   | 721 CCAAAATTCCTCTTGGAGTACATATCAAAAAGAAAGTTGGAGAGGCTTCCATAAAAAGTA 780   |
| Qy   | 261 ProGlnLeuLysMetGlnIle 267  |
| Db   | 781 CCACATTTAAAAATGAGATA 801   |

RESULT 3  
BC011904  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBMED  
REFERENCES  
AUTHORS  
TITLE  
JOURNAL  
REMARK  
COMMENT

BC011904 1908 bp mRNA linear PRI 23-DEC-2003  
Homo sapiens O-sialoglycoprotein endopeptidase-like 1, mRNA (CDNA  
clone MGC:20293 IMAGE:4121450), complete cds.  
BC011904  
BC011904.2 GI:40225818  
MGC.  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1908)  
Straussberg, R.L., Feingold, B.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, C.M., Schaefer, C.F., Bhat, N.K.,  
Altschul, S.F., Zeeberg, B., Bueltow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heien, F.,  
Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Schetz, T.E., Brownstein, M.J., Udell, T.B., Tohyuki, S.,  
Carinci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwen, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Ketterman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shcherchenko, Y.,  
Boultard, G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, U., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E.,  
Scherer, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16699-16903 (2002)  
12477932  
2 (bases 1 to 1908)  
Straussberg, R.  
Direct Submission  
Submitted (30-JUL-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
On Dec 19, 2003 this sequence version replaced gi:15080281.  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland.  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)  
Akhter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
Blakesley, R.W., Boultard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Granite, S., Guan, J., Gupta, C., Haghighi, P.,  
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R.,  
Madhus, O.L., Masello, C., Masker, B., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Stantirpop, S., Thomas, P.J., Touchman, J.W.,  
Tsurgenev, C., Vogt, J.L., Walker, M.A., Weherby, K.D., Wiggins, L.,  
Young, A., Zhang, L.-H. and Green, E.D.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRL plate: 28 Row: 1 Column: 22.  
Location/Qualifiers  
1. 1908  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:20293 IMAGE:4121450"  
/tissue\_type="Muscle, rhabdomyosarcoma"

gene  
cds  
misc\_feature  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,976-126 Length: 1908  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0  
US-10-649-273-2\_COPY\_148\_414 (1-267) x BC011904 (1-1908)  
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValAlaGluPheProPheLeuVal 20  
DB 545 ATGGAGGCTCAGCAGCTTACTTATTTGTTGACCAATTAAGTGAATTTCTTTTACTT 604  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
DB 605 CTTTGTGATTTCTGAGAGTCACTGCTGTTGGCATTAGTTCAAGAGTTCAGATTTTCG 664  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 60  
DB 665 CTTCTTGAAAGTCTTTGACATAGCACAGGTGACATGCTTGACAAAGTGCAGGAAGA 724  
QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 80  
DB 725 CTTCTTTTAATTAACATCCAGAGTCTCCACCAAGAGGTGGGAAAGCCATTAGACAT 784  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
DB 785 TTGGCCAAACAGAAATGATTTTCAATTTTGAATCAACCTCCCTTGATCATGCTAAA 844  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
DB 845 AATTGTGATTTTCTTTTACTGACTTCAACGCTTACGATTAATAATAATGAAAAAG 904  
QY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIle 140  
DB 905 GAAAAAGGAGGAAGTATGAGAAAGGCAAAATCTGTCTTCAGACAGACATTCCTGCC 964  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
DB 965 AAGAGTACGACACACATGATGATATCTTGAAAAAGAACATCGGCTATTTCTGTTT 1024  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180  
DB 1025 TGTAAAGCAGAGAGCTTTTACTCTCAAAATAATGACAGTACTGTTGCACTGCTGTGTGC 1084



QY 181 AAserspNheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200  
DB 1085 GGAAGTAACCTTCAATATCCGACAGCTCTGGAATTTTAAACCAACGACACAGTGCCT 1144  
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTyrAsnGly 220  
DB 1145 TTGTGTGTCTCTCCACAGCTATGACATGATTAATGAGCATTTGATGATGAGAAAGT 1204  
QY 221 IlleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 240  
DB 1205 ATTGAAAGACTACGTCTGCTGGCATTTTACATGACATGAGGACATCCGCTATGAA 1264  
QY 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260  
DB 1265 CCAAAATGTCTCTTGGAGTACATATCAAAAGAAAGTTGGAGAGCTTCATTAAGTA 1324  
QY 261 ProGlnLeuLysMetGluIle 267  
DB 1325 CCACAAATTAATAATGAGATA 1345  
RESULT 4  
LOCUS AR428803 2197 bp DNA linear PAT 18-DEC-2003  
DEFINITION Sequence 1 from patent US 6642041.  
ACCESSION AR428803  
VERSION AR428803.1 GI:40188589  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 2197)  
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.  
TITLE Polynucleotides encoding a novel metalloprotease, MP-1  
JOURNAL Patent: US 6642041-A 1 04-NOV-2003;  
FEATURES  
source Location/Qualifiers  
1..2197  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Alignment Scores:  
Pred. No.: 2,33e-126 Length: 2197  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x AR428803 (1-2197)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
DB 672 ATGAGAGCTCATGCACTTACTATTAGGTGACCAATTAAGTAGAATTCCTTTTGT 731  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
DB 732 CTTTGAATTTCTGAGAGTCACTGTCTGTGGCATTTAGTTCAAGAGTTTCAAGATTTCTG 791  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
DB 792 CTTCTTGGAAAGCTTTGACATAGCACAGTGACATGCTTGGCAAGGTGGCAAGAGA 851  
QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyValAlaIleGlnHis 80  
DB 852 CTTTCTTTAATAAATCAATCCAGAGTCTCCACATGAGTGTGGGAAGCCATAGAACAT 911  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisAspIleLysProPheLeuHisAlaLys 100  
DB 912 TTGGCCAAACAAAGAAATATGATTTCAATTGACATCAAACTCTCTTGCATCAGCTAAA 971  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
DB 972 AATTGATTTTCTTTACTGACATTCACACGTTTACTATATAATATATCAAAAAG 1031

QY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140  
DB 1032 GAAAGAGAGAGAGTATTTGAGAGAGGCAATCTGTCTTCAGCAGCAGCATTCCTCC 1091  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisAlaGalaIleLeuPhe 160  
DB 1092 ACAATACAGACACCAATGAGATGTCATCTGTGAAAAGAACACATCGGGCTAATCTGTT 1151  
QY 161 CysLysGlnArgAspIleLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 180  
DB 1152 TGTAGCAGAGAGACTGTGTAATCTCAAAATATACAGACTGTTGCAATCTGTGTGTTC 1211  
QY 181 AAserspNheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200  
DB 1212 GCAAGTAACCTTCAATATCCGACAGCTCTGGAATTTTAAACCAACGACACAGTGCCT 1271  
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTyrAsnGly 220  
DB 1272 TTGTGTGTCTCTCCACAGCTATGACATGATTAATGAGCATTTGATGATGAGAAAGT 1331  
QY 221 IlleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 240  
DB 1332 ATTGAAAGACTACGTCTGCTGGCATTTTACATGACATGAGGACATCCGCTATGAA 1391  
QY 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260  
DB 1392 CCAAAATGTCTCTTGGAGTACATATCAAAAGAAAGTTGGAGAGCTTCATTAAGTA 1451  
QY 261 ProGlnLeuLysMetGluIle 267  
DB 1452 CCACAAATTAATAATGAGATA 1472  
RESULT 5  
LOCUS AR428808 1387 bp DNA linear PAT 18-DEC-2003  
DEFINITION Sequence 21 from patent US 6642041.  
ACCESSION AR428808  
VERSION AR428808.1 GI:40188594  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 1387)  
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.  
TITLE Polynucleotides encoding a novel metalloprotease, MP-1  
JOURNAL Patent: US 6642041-A 21 04-NOV-2003;  
FEATURES  
source Location/Qualifiers  
1..1387  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Alignment Scores:  
Pred. No.: 5,43e-124 Length: 1387  
Score: 1358.50 Matches: 266  
Percent Similarity: 91.10% Conservative: 0  
Best Local Similarity: 91.10% Mismatches: 1  
Query Match: 98.09% Indels: 25  
Gaps: 1

US-10-649-273-2\_COPY\_148\_414 (1-267) x AR428808 (1-1387)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
DB 465 ATGAGAGCTCATGCACTTACTATTAGGTGACCAATTAAGTAGAATTCCTTTTGT 524  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
DB 525 CTTTGAATTTCTGAGAGTCACTGTCTGTGGCATTTAGTTCAAGAGTTTCAAGATTTCTG 584  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60

Db 585 CTTCTTGAAAGCTTTTGACATAGCACAGGTGACATGCTTGACAAAGTGGCAAGAA 644  
Qy LeuSerLeuIleuYshisProGluCySerTherMetSerGlyLyalaIleGluHis 80  
Db 645 CTTTCTTTAATAAACAATCCAGAGTGTCTCCACCATGAGTGTGGAAAGCATAGAGCAT 704  
Qy LeuAlaIysGlnGlyAsnArgPheHisPheAspIleuYsProPoleuHisAlaIlys 100  
Db 705 TTGGCCAAACAGAGAAATAGATTTCATTGTGACATCAAACTCCCTTGACATGCTAAA 764  
Qy 101 AsnCyAspPheSerPheThnGlyLeuGlnHisValThraPheYsIleIleMetIlylys 120  
Db 765 AATGTGATTTTCTTTTACTGACCTTCAACACGTTACTGATTAATAATAGAAAAAG 824  
Qy 121 GluIlysgIuGluGlyIle-----GluYsGlyGlnIleLeuSerSerAla 126  
Db 825 GAAAAAGAGAAAGTATATTTCTAATTAAGTGAACGATTAATAATTCCTGGATTG 884  
Qy 127 -----GluYsGlyGlnIleLeuSerSerAla 135  
Db 885 TGCCATAAATAAGTGTCTCATTTCTGACAGTATGAGAAAGGGCAATCTCTCTTCAACA 944  
Qy 136 AlaAspIleAlaAlaThraValGlnHisTherMetAlaCyshisIleuValIysArgThris 155  
Db 945 GCAGACATTTGCTGTCACAGTACACACAAATGGCATGTCTTGTGAAAAAGAACACAT 1004  
Qy 156 ArgAlaIleLeuPheCyshisGlnArgAspIleuProGlnAsnAsnAlaValIleuVal 175  
Db 1005 CGGGCTATTCTGTTTGTAGCAGAGAGACTGTGTACTCAAAATTAAGCAGTACTGTT 1064  
Qy 176 AlaSerGlyGlyValAlaSerAsnPheThyIleArgArgAlaIleGluIleLeuThrasn 195  
Db 1065 GCATCTGTGTGTGTGCGACAGTACTTCTATATCCGACAGCTCTGGAATTTTAAACAAC 1124  
Qy 196 AlaThrgIlnCyshisLeuCyshisProProArgLeuCyshisAspAspGlyIleMet 215  
Db 1125 GCAACACAGTGCATTTGTGTGTCTCTCCACAGACTATGACATGATTAATGGCATTAAG 1184  
Qy 216 ILeuAlaThraPheGlnIleGluArgLeuArgAlaGlyIleGluGlyIleLeuHisAspIleGlu 235  
Db 1185 ATTGCAATGGAATGTATTTGAAAGACTACGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1244  
Qy 236 GlyIleArgGlyArgProIlyCyshisProLeuGlyValAspIleSerIlyGlyValIlyGlu 255  
Db 1245 GGCATCCGCTATGAAACCAAAATGCTCTTGGAGTATGACATATCAAAAGAGTGGAGAA 1304  
Qy 256 AlaSerIleIysValProGlnLeuIlyMetGluIle 267  
Db 1305 CTTTCCATAAAGATACCAATTAATAAATGAGAGATA 1340  
RESULT 6  
LOCUS HSA295148 1387 bp mRNA linear PRI 30-OCT-2000  
DEFINITION Homo sapiens mRNA for putative sialoglycoprotease type 2.  
ACCESSION AJ295148  
VERSION AJ295148.1 GI:11071726  
KEYWORDS metalloproteinase; sialoglycoprotease.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1  
AUTHORS Chen, J.M., Fortunato, M. and Barrett, A.J.  
TITLE Cloning and sequencing of a second human putative  
sialoglycoprotease homologue  
JOURNAL Unpublished  
AUTHORS Chen, J.M.  
TITLE Direct Submission  
JOURNAL Submitted (27-OCT-2000) Chen J.M., MRC Molecular Enzymology  
Laboratory, The Babraham Institute, Babraham, Cambridge, CB2 4AT,  
UNITED KINGDOM  
FEATURES  
Location/Qualifiers

source 1..1387  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/tissue="uterus"  
24..1343  
/function="metalloproteinase of family M22"  
/codon\_start=1  
/product="putative sialoglycoprotease type 2"  
/protein\_id="CAC14666.1"  
/db\_xref="GI:11071727"  
/db\_xref="GOA:O9H4B0"  
/db\_xref="UniProt/TREMBL:O9H4B0"  
/translation="MLILTKTAGVFPKSRVVEFLRSFNFPELTLPLAKIVLGIET  
SCDDTAAAVVDEGTGVLGEALHSQTEVHLTKGVIYPAPQQLHRENIQRIQVEALAS  
GVSPSDLSAIVATTIKPGLALSLGVLSFSLQVGLKKPFIPIHMEHALATIRLTK  
VEPPIVILLISGGHCLIALVGVSPDLISGLSDIAPDMDLKVRLSLIKHPECT  
MSGKALIEHLAKQGRFHPDIKPLIHLAKNCDPSTGLOHTYDILMKKEKEGFLI  
SKVQINIPGLCLKIAHFCREREKQILSSADIAATVQHTACHLVKRTTHAALIFCK  
ORDLLPQNNAVILVAGVANSFYIRRALEILTNATQCLCPPLCTDNGIMIAMNG  
IERLRGIGILHIDIEGIRYEPKPLGVNIDISKEVBSIKVPLKMEI"  
ORIGIN  
Alignment Scores:  
Pred. No.: 5,43e-124 Length: 1387  
Score: 1358..50 Matches: 266  
Percent Similarity: 91.10% Conservative: 0  
Best Local Similarity: 91.10% Mismatches: 1  
Query Match: 98.09% Indels: 25  
DB: 9 Gaps: 1  
US-10-649-273-2\_copy\_148\_414 (1-267) x HSA295148 (1-1387)  
Qy 1 MetGluAlaHisAlaLeuThriIleArgLeuLeuThraAsnIysValGluPheProPheLeuVal 20  
Db 465 ATGAGGCTCATGACCTTACTATAGGTTAGCAATTAAGTGAATTCCTTTTATGTT 524  
Qy 21 LeuLeuIleSerGlyGlyHisCyshisLeuAlaIleValGlnIlyValSerAspPheLeu 40  
Db 525 CTTTGTATTTCTGAGAGTCACTGTCTGTGGCATTAAGTTCAAGAGCTTCAAGATTTTCTG 584  
Qy 41 LeuLeuGlyIlySerIleuAspIleAlaProGlyAspMetLeuAspIysValIlaArgArg 60  
Db 585 CTTCTTGAAAGTCTTTTGGACATACACAGGTGACATGCTTGACAGGTGGCAAGAA 644  
Qy 61 LeuSerLeuIleYshisProGluCySerTherMetSerGlyLyalaIleGluHis 80  
Db 645 CTTTCTTTAATAAACAATCCAGAGTGTCTCCACCATGAGTGTGGAAAGCATAGAGCAT 704  
Qy 81 LeuAlaIysGlnGlyAsnArgPheHisPheAspIleuYsProPoleuHisAlaIlys 100  
Db 705 TTGGCCAAACAGAGAAATAGATTTCATTGTGACATCAAACTCCCTTGACATGCTAAA 764  
Qy 101 AsnCyAspPheSerPheThnGlyLeuGlnHisValThraPheYsIleIleMetIlylys 120  
Db 765 AATGTGATTTTCTTTTACTGACCTTCAACACGTTACTGATTAATAATAGAAAAAG 824  
Qy 121 GluIlysgIuGluGlyIle-----GluYsGlyGlnIleLeuSerSerAla 126  
Db 825 GAAAAAGAGAAAGTATATTTCTAATTAAGTGAACGATTAATAATTCCTGGATTG 884  
Qy 127 -----GluYsGlyGlnIleLeuSerSerAla 135  
Db 885 TGCCATAAATAAGTGTCTCATTTCTGACAGTATGAGAAAGGGCAATCTCTCTTCAACA 944  
Qy 136 AlaAspIleAlaAlaThraValGlnHisTherMetAlaCyshisIleuValIysArgThris 155  
Db 945 GCAGACATTTGCTGTCACAGTACACACAAATGGCATGTCTTGTGAAAAAGAACACAT 1004  
Qy 156 ArgAlaIleLeuPheCyshisGlnArgAspIleuProGlnAsnAsnAlaValIleuVal 175  
Db 1005 CGGGCTATTCTGTTTGTAGCAGAGAGACTGTGTACTCAAAATTAAGCAGTACTGTT 1064

QY 176 AlaSerGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeuThrAsn 195  
DB 1065 GCATGCGTGTGTCCAGACTAATCTTATATCCGACAGCTCTGGAAATTTTAAACAAC 1124  
QY 196 AlaThrGlnCysThrIleuLeuGlyProProArgIleuGlySerIleAspAsnGlyIleMet 215  
DB 1125 GCAACACAGTGCACCTTGTGTGTGCTCCCTCCAGACTATGACATGTAATATGGCATATATG 1184  
QY 216 IleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlu 235  
DB 1185 ATTCGATGGAATGATGATGGAAGACATACGTCGTGGCTTGGCATTTACATGACATGAA 1244  
QY 236 GlyIleArgTyrGluProIleGlyCysProLeuGlyValAspIleSerIleGluValGlyGlu 255  
DB 1245 GGCATCCGCTATGAAACCAAAATGCTCCTTGAGTACATATCAAAAGAAGTTGAGAA 1304  
QY 256 AlaSerIleLeuValProGlnLeuIleuMetGluIle 267  
DB 1305 GCTTCATAAAGTACCAATTAATAATGAGATA 1340  
RESULT 7  
AX664697 1245 bp DNA linear PAT 22-MAR-2003  
LOCUS Sequence 6 from Patent WO02074960.  
ACCESSION AX664697  
VERSION AX664697.1 GI:29164457  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS 1  
TITLE Leiby, K.R., Kapeller-Libermann, R. and Glucksmann, M.  
JOURNAL 38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions  
of human proteins and uses thereof  
Patent: WO 02074960-A 6 26-SEP-2002;  
Millennium Pharmaceuticals, Inc. (US)  
FEATURES  
source 1.1245  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Alignment Scores:  
Pred. No.: 5.14e-123 Length: 1245  
Score: 1348.00 Matches: 260  
Percent Similarity: 98.50% Conservative: 3  
Best Local Similarity: 97.38% Mismatches: 4  
Query Match: 97.33% Indels: 0  
DB: 6 Gaps: 0  
US-10-649-273-2\_COPY\_148\_414 (1-267) x AX664697 (1-1245)  
QY 1 MetGluAlaHisIleAlaLeuThrIleArgLeuThrAsnIleValGluPheProPheLeuVal 20  
DB 442 ATGGAGGCTCATGCACTTACTATTAAGTTACCAATAAGTAGAATTTCTTTTATGTT 501  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
DB 502 CTTTTCATTTCTTGAGAGTCACTGCTGTGTGGCATTAAGAGATTTCAGATTTTCTG 561  
QY 41 LeuLeuGlyIleSerLeuAspIleAlaProGlyIleAspMetLeuAspIleValIleArgArg 60  
DB 562 CTTTCTGGAAGCTTTTGACATAGCAACGAGTGACATGCTTGACAGAGTGGCAAGAAGA 621  
QY 61 LeuSerIleIleHisIleProGluCysSerThrMetSerGlyGlyValIleGluHis 80  
DB 622 CTTTCTTAATTAATAACATTCAGAGTGTCTCAACATGATGATGAGAGGCAATGAAAT 681  
QY 81 LeuAlaIleGlnGlyAsnArgPheHisIleAspIleIleValProLeuHisIleIleVal 100  
DB 682 TTGGCAAAACAAGGAATATGATTTTCATTTTGCATCAACCTCTTGCATCATGCTAAA 741

QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIleIleMetLeuLys 120  
DB 742 AATGTGATTTTCTTTTACTGAGACTCAACACGCTTACTGTATTAATAATGAAAAACAG 801  
QY 121 GluIleGluIleGlyIleGlyIleGlyGlnIleLeuSerSerAlaIleAspIleAla 140  
DB 802 AAACAAAGAGAAAGTATTTGAAAGGGGCAATCTGTCTTACAGACAGCATTTGCTGCC 861  
QY 141 ThrValGlnHisIleThrMetAlaCysHisIleuValIleAspArgThrHisIleValIle 160  
DB 862 ACAGTACAGACACAAAGGACATGTCATCTTGTGAAAGAAACATCGGCTATTTCTGTT 921  
QY 161 CysIleGlnArgAspLeuLeuProGlnAsnAlaValLeuValAlaSerGlyVal 180  
DB 922 TGTATGACAGAGAGACTTGTATACCTCAAAATATATCAGATGCTGTTGATCTGTGTC 981  
QY 181 AlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200  
DB 982 GCAAGTAACTTCTATATCCGACAGCTCTGGAAATTTTAAACAAACGCAACACAGTGCCT 1041  
QY 201 LeuLeuGlyProProArgLeuGlyCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220  
DB 1042 TTGTGTGCTCTCTCCAGACTATGACATGATATGATGATGATGATGATGATGATGAT 1101  
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 240  
DB 1102 ATTGAAGACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1161  
QY 241 ProIleGlyCysProLeuGlyValAspIleSerIleGlyValGlyGluAlaSerIleVal 260  
DB 1162 CCAAAATGCTCTCTTGAGTACATATCAAAAGAAGTTGAGAGAGCTTCCATTAATAAGTA 1221  
QY 261 ProGlnLeuIleuMetGluIle 267  
DB 1222 CCACAAATTAATAAATGAGAGATA 1242  
RESULT 8  
AX664695 1820 bp DNA linear PAT 22-MAR-2003  
LOCUS Sequence 4 from Patent WO02074960.  
ACCESSION AX664695  
VERSION AX664695.1 GI:29164455  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS 1  
TITLE Leiby, K.R., Kapeller-Libermann, R. and Glucksmann, M.  
JOURNAL 38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions  
of human proteins and uses thereof  
Patent: WO 02074960-A 4 26-SEP-2002;  
Millennium Pharmaceuticals, Inc. (US)  
FEATURES  
source 1.1820  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
146.1390  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD80044.1"  
/db\_xref="GI:29164456"  
/translation="MLILRTTAVGVFFKRSKKRYEFLRSFNFHPTLFLHKIVLGIEIT  
SCDDTAAVVDENGVGBAIIHQTEVHLTKGIVPPAAQQIARENIQRIYQVBLAS  
GVSPSDSAVATITKPGALSLGVGLSFSLOVOLKKEPPIPHHMEHALITRLTK  
VEPPIVLLISGGCILALVOGSDPFLLSKSLIABQMDKVARRLSLIKHPEGST  
MSGKALIEHAKQENRPHDIKPIHLIAKNCDEPFGIQTDTDNENRKBEGIEG  
QILSSAADIAVQHTWACHLVKTRHAIILFCQRDLIPONNAVIVASGVASNPYIR  
RALEILTNATQCTLCPPRLCTDNGIMAWNGIERLRAGIGIILHDIIGIIEBXCPL  
GVDSKEVGASIKVPLKMEI"

ORIGIN



```

CDS
    /note="Vector: PCMV-SPORT6"
    222..1466
    /codon_start=1
    /product="unknown (protein for MGC:67870)"
    /protein_id="AAH58172.1"
    /db_xref="GI:34849664"
    /translation="MTMRTTGAIPKPPSKYKYGIRRSVHPRTLSCHKVLGIRT
    SCDDTAAVVDENGVNGLSHQTVHLEGTGIVPPVAQVLRHNIQRIVEETLSAS
    RITPDLISAATTIKPGLAISLGVLSFSLQVNRPKKPIPIHMEAHALITLTKK
    VEPFPLVLISGGHCLALVQGVDFLLGKSLIDLAGMDLKVARSLSIKPECSY
    MSGGKAIEHLAKKDNFHTINPMDNAKKDPSFTGLQHTITVLTKHKEGIEKG
    OLISSADIIAAVQHAATACILAKRTERRALIFCKQNLSPANVILKLVKSNLYR
    KALEIVANROCTLLCPPRCLCTDNGIMIANNGIERRRAGLGVLDVEDIRYKPCPL
    GIDISREVAEAAIKVPLKVAL"
    misc_feature
    333..1397
    /note="ORF17: Region: Metal-dependent proteases with
    possible chaperone activity [posttranslational
    modification, protein turnover, chaperones]"
    /db_xref="CCD:COG0533"

ORIGIN
Alignment Scores:
Pred. No.:      2,76e-111      length:      1844
Score:          1231.00      Matches:      233
Percent Similarity: 92.88%      Conservative: 15
Best Local Similarity: 87.27%      Mismatches: 19
Query Match:      88.88%      Indels:      0
DB:              10      Gaps:      0

US-10-649-273-2_COPY_148_414 (1-267) x BC058172 (1-1844)
QY      1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB      663 ATGGAGGCGTCACGACCTGATTAAGGCTCACCAATAGATTTCTTTCTTCTAGTT 722
QY      21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB      723 CTTTTCATTTCTTGCGCGTCACTCCCTGTTGGCATTAGTCCAGGCTTTCCGATTCTCG 782
QY      41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB      783 CTCCTTGGGAGGCTTTGGACATAGCGCCAGCGACATGCTTGAACAAGGTGGCAAGAGA 842
QY      61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyValAlaIleGlnHis 80
DB      843 CTTTCTTATCAACATCCAGATGTTCTTACATAGTGTGGAAGACTATAGAACAT 902
QY      81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
DB      903 TTGGCCAAAGACCGAATAGATTCATTTTACTATCAATCCACTATGACAAATGCTAAG 962
QY      101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
DB      963 AATTCGATTTTCTTTTACGGGACCTTCAACATATTACTATTAAGCTAAATTAACACACAG 1022
QY      121 GlnLysGlnGluGlyIleGlyLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
DB      1023 GAAAAAGAAAGAGGCACTTGAAGGGGCAATCTCTGTCATCACTGACAGACATTTGCTGCT 1082
QY      141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
DB      1083 GCGGTACAGATCAACAGCGGTCCACCTTGCAAAAAGAACACATCGTGCATATCTGTTT 1142
QY      161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
DB      1143 TCCAAACCAAAAAAATTTGCTATCTCCAGCTACGACTATTAAGTTGATCTGAGGGGTGT 1202
QY      181 AlaSerAspPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
DB      1203 GGAAGTAACTGTGATCCGAAAAAGCAATGGAATTTGTCGCAATATGCAAGCGAGTGACG 1262
QY      201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220

```

```

DB      1263 TTGCTGTGTCCTCCCTCCAAAGACTGTGACATGACATGACATGATTCATGCAATGCA 1322
QY      221 IlleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlyGlyIleArgTyrGlu 240
DB      1323 ATTCAAAGATTAAGTCGTGGCTGGCTTTTACATGATGATGAGAACATCGATATATCA 1382
QY      241 ProLysCysProLeuGlyValAlaAspIleSerLysGlyValGlyGlnAlaSerIleLysVal 260
DB      1383 CCAAAATGTCCTCTTGGAAATGACATATTCAGAGAAAGTTGACAGAGCTGCATTAAGTA 1442
QY      261 ProGlnLeuLysMetGluIle 267
DB      1443 CCGGATTAATAATGCACTT 1463

RESULT 10
BC038910
LOCUS
DEFINITION Mus musculus O-6-ialoglycoprotein endopeptidase-like 1, mRNA (cDNA
clone IMAGE:503559), partial cds.
ACCESSION BC038910
VERSION BC038910
KEYWORDS BC038910.1 GI:24433548
SOURCE
ORGANISM Mus musculus (house mouse)
MUS musculus; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1017)
REFERENCE
AUTHORS Klausner,R.D., Collins,F.S., Wagner,L., Sherman,C.M., Schuller,G.D.,
Altschul,S.F., Zeeberg,B., Buettow,K.H., Schaefer,C.F., Bhac,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heideh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schoeetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loggellano,N.A., Peters,G.J.,
Adamsom,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahy,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalls,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marr,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
2388257
JOURNAL MEDLINE
PUBMED 12477932
REFERENCE
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (25-OCT-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK
COMMENT NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-rc@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasha van den Bosch, Jill Vardy,

```

George Yang, Scott Zuyderduyn, Marco Marra

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Series: IRAX Plate: 86 Row: F Column: 12.

Location/Qualifiers  
1. .1017

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="PVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5053559"
/risuse_type="river, normal, 5 month old male mouse."
/clone_id="NCI CGAP_119"
/lab_host="DH10B"
/lab_note="PCMV-SPORT6"
note="vector"

```

ORIGIN

Alignment Scores:

|                        |           |                  |
|------------------------|-----------|------------------|
| Pred. No.:             | 6.53e-111 | 101              |
| Score:                 | 1224.00   | 233              |
| Percent Similarity:    | 92.51%    | Conservative: 14 |
| Best Local Similarity: | 87.27%    | Mismatches: 20   |
| Query Match:           | 88.38%    | Indels: 0        |
| DB:                    | 10        | Gaps: 0          |

US-10-649-273-2\_COPY\_148\_414 (1-267) X BC038910 (1-1017)

|    |  |     |
|----|--|-----|
| Qy | MetGluAlaHisAlaIleThrIleArgLeuThrAsnLysValGlnPheProPheLeuVal | 20  |
| Db | ATGAGAGCTCAGCGACTGACTATTAAGGCTCACCAATAAAGTAGAATTTCTTTTAGTT   | 85  |
| Qy | LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu | 40  |
| Db | CTTTTGATTTCTCGCGGTCACTGCCTGTGGCATTTGTCCAAAGGTGTTCCGATTTCTTG  | 145 |
| Qy | LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg | 60  |
| Db | CTCCTTGGAAGTCTTTGGACATAGCACAGGCGACATGCTTGACAAAGGTGGCAAGAAGA  | 205 |
| Qy | LeuSerLeuIleLysHisAspGlyCysSerThrMetSerGlyLysAlaIleGlnHis    | 80  |
| Db | CTTTCTTTAATCAAAATCATCCAGATGTTCTACAACTAGTGGTGGAAAGCTATAGAACAG | 265 |
| Qy | LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProPheLeuHisAlaLys    | 100 |
| Db | TTGGCGAAAGCGGAAATAGATTTCCATTTACTATCAATCCACTATGCAAGGCTTAAG    | 325 |
| Qy | AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys | 120 |
| Db | AATTCGATTTTCTTTCAAGGGACTTCAACATATTACTGATAGCTAATTAACACACAG    | 385 |
| Qy | GlyLysGlnGlnGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla | 140 |
| Db | GAAAAAGAAAGAGCATTGAGAAAGGGGCAAAATCTGTCACTACGCTCAGACATTCGCT   | 445 |
| Qy | ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe | 160 |
| Db | GCGGTACAGCAGCAACAGCGTCCACTTGCAGAAAGAACACATGCGCTATTCTGTTT     | 505 |
| Qy | CysLysGlnArgAspLeuLeuProGlnIleAsnAsnAlaValLeuValAlaSerGlyVal | 180 |
| Db | TGCAGACGAAAAATTTGCTCTCTCCAGCTAAAGCAGTATTAGTTGATTCGAGGGTT     | 565 |
| Qy | AlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr | 200 |
| Db | GCAATTAATCTGTACATCCGAAAGATTTGGAATTTGCGCAATGCAAGCAGTGCAGC     | 625 |
| Qy | LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTyrAsnGly | 220 |
| Db | TTGTTGTGTCCACTTCGAAGACTGTGCATCGACAAATGCATCAATGATGATGAATGGA   | 685 |
| Qy | IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGln | 240 |

Db 686 ATTGAAGATTACGTGCTGGCTTGGCGTTTACATGATGTAGAAGACATCCGATATGAA 745

|    |     |   |     |
|----|-----|---|-----|
| Qy | 241 | ProLysCysProLeuGlyValAspIleSerIleGluValGlyIuAlaSerIleIysVal | 260 |
| Db | 746 | CCAAATGTCCTCTTGAGTAGACATATCCAGAGAAGTTCGACAGAGCTGCCATTAAGTA  | 805 |

```
QY      261 ProGlnLeuIysMetGluIle 267
      |||:::|||||
Db      806 CCGGATTAAAAATGGCACTT 826
```

## RESULT 11

|            |                                     |            |     |        |                 |
|------------|-------------------------------------|------------|-----|--------|-----------------|
| LOCUS      | AX13716                             | 2208 bp    | DNA | linear | PAT 15-APR-2003 |
| DEFINITION | Sequence 400 from Patent EP1293569. |            |     |        |                 |
| ACCESSION  | AX13716                             |            |     |        |                 |
| VERSION    | AX13716.1                           | GI:2988642 |     |        |                 |

**SOURCE**

| SOURCE  | ORGANISM     |
|---|--------------|
| Homo sapiens (human)  | Homo sapiens |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. |              |

## REFERENCES

1  
I  
Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,  
Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,  
Tanechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and  
Masuko, Y.

## JOURNAL

Helix Research Institute (JP) ; Research Association for

## FEATURES

```
source 1. .2208
        /organism="Homo sapiens"
        /mol_type="unassigned DNA"
        /db_xref="taxon:9606"
```

## ORIGIN

**Alignment Scores:**

|                        |           |               |     |
|------------------------|-----------|---------------|-----|
| Pred. No.:             | 1,576-108 | Length:       | 230 |
| Score:                 | 1304.00   | Matches:      | 229 |
| Percent Similarity:    | 89.51%    | Conservative: | 0   |
| Best Local Similarity: | 89.51%    | Mismatches:   | 4   |
| Query Match:           | 86.93%    | Indels:       | 24  |
| DB:                    | 6         | Gaps:         | 1   |

US-10-649-273-2\_COPY\_148\_414 (1-267) X AX713716 (1-2208)

|    |      |   |      |
|----|------|---|------|
| Qy | 1    | MetGIuAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal    | 20   |
| Db | 785  | ATGGAGGCTCAGACCTTAATAAGTGTGACCAATAAAGTAATAATTTCTTTTAAAGTT       | 844  |
| Qy | 21   | LeuLeuIleSerGluGlyVhiScYsLeuLeuAlaLeuValGlnGlyValSerAspPheLeu   | 40   |
| Db | 845  | CTTTTGATTTCTGAGAGTCACTGTCTGTGGCATTAAGTTCAAGAGATTTCAGATTTTCTG    | 904  |
| Qy | 41   | LeuLeuGluLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg    | 60   |
| Db | 905  | CTTCGTGAAAGTCTTTGGACATATACACACAGGAGCATCTTGACAAAGGTGCAGAGAGA     | 964  |
| Qy | 61   | LeuSerLeuIleLysHisProGluGlySerThrMetSerGluGlyLysAlaIleGluHis    | 80   |
| Db | 965  | CTTCCTTTAAATAAACATCCAGAGGCTCCACCATAGTGGTGGAAAGCCATAGAACAT       | 1024 |
| Qy | 81   | LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisAlaLys       | 100  |
| Db | 1025 | TTGGCCAAACAGAGAAATAGATTTCATTTTGACATCAAACTCCCTTGACATATCTTAA      | 1084 |
| Qy | 101  | AsnGlyAspPheSerPheThrGlyLeuGlnHisCysAlaHisAspLysIleIleMetLysLys | 120  |
| Db | 1085 | AATGTGATTTTCTCTTTTAAGTGAACCTTCAACAGCTTACGATTAATAATTAATGAAG      | 1144 |
| Qy | 121  | GluLysGluGluGlyTlreGluLysGluGlnIleLeuSerSerAlaAlaAspIleAla      | 140  |

| Db         | 1145  | GAATAAGGAGAGGATTGAGAGGGGAAATTCCTGCTTCGACGACGACGACTTCCTCC           | 1200 |
|------------|---|--|------|
| Qy         | 141   | ThrValGlnHisThrMetAlaCysHisLeuValIysArgThrHisArgAlaIleLeuPhe       | 160  |
| Db         | 1205  | ACAGTACGACGACACACATGCGATGTCATCTGTTGAAAAAACAACATCGGGCTATTCGTGTT     | 1264 |
| Qy         | 161   | CysLeysGlnArgAspLeuLeuProGlnIAsnAsnAlaValIleuValAlaIleSerGlyGlyVal | 180  |
| Db         | 1265  | TGTAAAGCAGAGAACTGTGTACTTCCTCAAAATATATGACGATCGGTGCAATCGTGAGTGTCTC   | 1324 |
| Qy         | 181   | AlaSerAsnPhePheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr    | 200  |
| Db         | 1325  | GCAAGTAACTCTGATATCCGACGACTTCGAAATTTTAAACAACGACACAGTCGACT           | 1384 |
| Qy         | 201   | LeuLeuCysProProProArgLeuLeuCysThrAspGlnGlyIleMetIleAlaATPAsnGly    | 220  |
| Db         | 1385  | TTGTGTGTCCTCTCCCTCCAGACTATGCACTGATTAATGCAATATGATTCGA-----          | 1435 |
| Qy         | 221   | IleGluArgLeuAlaGlyIleLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu       | 240  |
| Db         | 1435  | -----  | 1435 |
| Qy         | 241   | ProLysCysProLeuGlyValAspIleSerIysGluValGlyGlyAlaSerIleLysVal       | 260  |
| Db         | 1436  | ---TGATGTCCTCTTGGAGTGAATATCAAAAGAGTTGGAGAGCTTCCTCAATAAGTA          | 1492 |
| Qy         | 261   | ProGlnLeuLysMetGluIle  | 267  |
| Db         | 1493  | CCACACATTTAAATATGACAGATA   | 1513 |
| RESULT 12  |   |  |      |
| LOCUS      | AK055441  |  |      |
| DEFINITION | Homo sapiens cDNA FLJ30879 fis, clone FBRB20049529, highly similar to Homo sapiens mRNA for putative staloglycoprotease type 2.   |  |      |
| ACCESSION  | AK055441  | GI:16550166  |      |
| VERSION    | AK055441.1  | GI:16550166  |      |
| KEYWORDS   | oligo capping; fis (full insert sequence).  |  |      |
| SOURCE     | Homo sapiens (human)  |  |      |
| ORGANISM   | Homo sapiens  |  |      |
| REFERENCE  |   |  |      |
| AUTHORS    | 1 Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Ohayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Itoh, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M., Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimura, M., Watanabe, M., Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Hosiida, M., Hoshida, T., Kusano, J., Kanehori, K., Takahashi, Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togiyasu, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Mueshino, K., Yunki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishiguchi, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kunikida, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujizara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hiro, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N., Nishigaki, H., Ikema, Y., Okamoto, S., Ohtani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, T., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togaishi, T., Oyama, M., Hara, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. |  |      |

|  |   |               |      |  |
|--|---|---------------|------|--|
| TITLE  | Complete sequencing and characterization of 21,243 full-length human cDNAs  |               |      |  |
| JOURNAL  | Nat. Genet. 36 (1), 40-45 (2004)  |               |      |  |
| REFERENCE  | 14702039  |               |      |  |
| AUTHORS  | 2<br>Niishi,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuna,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Maehuo,Y., Nagai,K. and Isogai,T.               |               |      |  |
| TITLE  | NEO human cDNA sequencing project   |               |      |  |
| JOURNAL  | Unpublished   |               |      |  |
| REFERENCE  | 3 (bases 1 to 2208)   |               |      |  |
| AUTHORS  | Isogai,T., Otsuki,T. and Sugiyama,T.  |               |      |  |
| TITLE  | Direct Submission   |               |      |  |
| JOURNAL  | Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomicseshi.co.jp; Tel:81-438-52-3975, Fax:81-438-52-3986)   |               |      |  |
| COMMENT  | NEO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI. |               |      |  |
| FEATURES   | Location/Qualifiers   |               |      |  |
| SOURCE   | 1..2208   |               |      |  |
|  | /organism="Homo sapiens"  |               |      |  |
|  | /mol_type="mRNA"  |               |      |  |
|  | /db_xref="taxon:9606"   |               |      |  |
|  | /clone="FEBRA2004592"   |               |      |  |
|  | /tissue_type="Brain"  |               |      |  |
|  | /clone_id="FEBRA2"  |               |      |  |
|  | /dev_stage="fetus"  |               |      |  |
|  | /note="cloning vector: pME18SFLJ3"  |               |      |  |
| CDS  | 344..1438   |               |      |  |
|  | /note="unnamed protein product"   |               |      |  |
|  | /codon_start=1  |               |      |  |
|  | /protein_id="BAB70923.1"  |               |      |  |
|  | /db_xref="GI:16550167"  |               |      |  |
|  | /translation="MLILTKRGVFEKSKRVLEFSLFNFGTLPLHKIVLIE<br>SCDDIAAVVDENGLDEAHSQTEVHLKTGIVLPAAQOLHRENIQIVDEALSA<br>GSSSDLSAIAITTKIPGLALSIGVLSLSQLQVQLKKEPIIHMEALITRLTN<br>VEPRLVILISGGHIALVQGVSDPFLIGKSDILAPGMDLKVARLPLIKHPECS<br>MSGKATLHLAKQGNRHFPIDKPLHLNAKNCODPPTGLOHYATDKIIMKEKEGEIEK<br>QILSSADIIAATVQHTMACHLVKRTHRIILFEKQDLPQNNAVLVASGVASNFCTI<br>RALEILTNAVQCTLLCPPPRLCTDNGMIA"   |               |      |  |
| ORIGIN   |   |               |      |  |
| Alignment Scores:  |   |               |      |  |
| Pred. No.:   | 1,576-108   | Length:       | 2208 |  |
| Score:   | 1204.00   | Matches:      | 229  |  |
| Percent Similarity:                                      | 89.51%  | Conservative: | 0    |  |
| Best local Similarity:                                   | 89.51%  | Mismatches:   | 4    |  |
| Query Match:   | 86.93%  | Indels:       | 24   |  |
| DB:  | 9   | Gaps:         | 1    |  |
| US-10-649-273-2_COPY_148_414 (1-267) x AK055441 (1-2208) |   |               |      |  |
| OY   | 1 MetGuaLaHsSaLaLeuThrIleArgLeuThrAsnLysValGIuPheProPheLeuVal 20  |               |      |  |
| Db   | 785 ATGAGAGGCTACGACTTACTATTAGGTGACCAATAAGTAGAATTTCCTTTTACGTT 844  |               |      |  |
| OY   | 21 LeuLeuIleSerGIyLHisCysLeuLeuAlaLeuValGIuGLyValSerAspPheLeu 40  |               |      |  |
| Db   | 845 CTTTGGATTTCGAGGCTCACTGCTGTTGGCATTAGTTCACAGGAGTTTCAGATTTCG 904   |               |      |  |
| OY   | 41 LeuLeuGLyLysSerLeuAspIleAlaProGLyAspMetLeuAspLysValAlaArgArg 60  |               |      |  |
| Db   | 905 CTTCTGGAAAGCTTTTGGACATACACACAGGTGACATGCTTGGACAAAGGTGGCAAGAAG 964  |               |      |  |



```

QY 61 leuSerleuIlelyshisProglucySserThrmeserIglylysalalegluhis 80
DB 965 CTTCCCTTATATAAATCATCCAGGTGCTCCACCATGATGCTGGGAAGCATGAAACAT 1024
QY 81 leuAlaIySGInglysanaIrgphEhisPhaspIlelySProPoleuhIshIAlaIlys 100
DB 1025 TTGGCCAAACAAAGGAATATGATTTTCAATTTTGACATCAAACTCCCTTGACATCACTGAAA 1084
QY 101 AsnCyAspPheSerPheThrGlyleuGlnHisIvalThrAspIlyIleIleMetIlylys 120
DB 1085 AATTGATTTTCTCTTTACTGACCTTCAACAGCTTACGTATATAAATAATATATAAAG 1144
QY 121 GluIySGInglylyIleGlylySGIlyGlnIleleuSerSeraIaIAspIleAlaI 140
DB 1145 GAAAAAGAGAGATTTGAGAAAGGGCAATCTGCTTCAGACAGACATCTGCTGCC 1204
QY 141 ThrValGlnHisThrMetAlaCyHisIleuValIysArgThrHisArgAlaIleleuPhe 160
DB 1205 ACGTACAGACACAAATGGCATGCTCTGTGAAAAAACAACATCGGGCTATTCTGTTT 1264
QY 161 CysIySGInArgAspIleuPleuProGlnAspAsnAlaIleuValAlaIAserIglylyVal 180
DB 1265 TGTAAACAGAGACATTTGTAACCTCAAAATAATGACATGCTGATCTGTGATGTC 1324
QY 181 AlaSerAspPheTyrlleArgArgAlaIleuGlnIleleuThrAsnAlaIThrGlnCysThr 200
DB 1325 GCAAGTAACCTTCTGTATCCGACAGCTCTGTGAAATTTTAAACAAAGCAACAGATGCACT 1384
QY 201 leuIleuCySProPProPProAArgIleuCySThrAspAsnGlyIleMetIleAlaIArgAsnGly 220
DB 1385 TTCTTGTGCTCTCTCCACAGCTATGACATGATATGCACTTATGATTTGCA----- 1435
QY 221 IleGluArgIleuArgAlaGlyIleuGlyIleleuHisaspIleGlyIleArgTyGlu 240
DB 1435 ----- 1435
QY 241 ProIySGIleuProIySGIleuValAspIleSerIySGIleuValGlyIleuValAserIleVal 260
DB 1436 ---TGAATGCTCTCTGTGAGTAAGCATATCAAAAGAGCTTGAGAGCTTCATATAAAGTA 1492
QY 261 ProGlnIleuIySGIleuIle 267
DB 1493 CCAAAATTAATAATGAGATA 1513

RESULT 13
LOCUS BC078974 1546 bp mRNA linear ROD 03-AUG-2004
DEFINITION Rattus norvegicus cDNA clone IMAGE:7111906, partial cds.
ACCESSION BC078974
VERSION BC078974.1 GI:50926879
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1546)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Datchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Bosak,S.A., McEwan,P.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulvik,S.W.,
Vallalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahney,J., Heitton,E., Kettelman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,B.D.,

```

```

JOURNAL Buttefield,Y.S., Krzywinski,M.I., Skalska,U., Smalins,D.E.,
REFERENCE Scherch,A., Schein,J.B., Jones,S.J., and Marra,M.A.
AUTHORS Generation and initial analysis of more than 15,000 full-length
TITLE human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED 12477932
2 (bases 1 to 1546)
DIRECTOR Director MGC Project.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2004) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Howard Jacobs
CDNA Library Preparation: Express Genomics
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNLT)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www.shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNLT at: http://image.llnl.gov
Series: IRAX Plate: 182 Row: F Column: 6
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein.
FEATURES
source
1..1546
Location/Qualifiers
1..1546
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:7111906"
/tissue_type="Testis, rat (Brown Norway)"
/clone_id="NIH_MGC_237"
/lab_host="DH10B"
/note="vector: pExpress1"
144..>1546
/codon_start=1
/product="Unknown (protein for IMAGE:7111906)"
/protein_id="AAH78974.1"
/db_xref="GI:50926880"
/translation="MLMSKTAGAIIPRPSSVNGFIRPNVOPRALFHHKLVIGIET
SCDPTAAAVDTEGNGVGEALHSOTENVLTGGLVYPPVAQQLHENTORIVBEALAS
GSPSDLSAIVTTIKRGLASLVGELSVSVLVNOFKPFIPIHMEHMAITRLTLTK
VGFPFLVLLISGHCCLALVOSVSDFLGLSLDLPADMDLKVARRSLIKHPECT
MSGKALIEHLAKGNRFHTINPWNANKCDFSTGLOHTVDKLIITKEKEBIEKG
OILSSADIDIAAVOHATACHLAKRTHRIILFCQGNLSPVNAVYVGVASNNIYIR
PALEIVANAQCTTLCPPRLCTONGIWIANGIERLRAGLGLHDVDIRPAPGS
ITSVLRVITATLTDLSVSTHYTHVHSLNGSRKANIQTSMSCSTQTVVMTVHTLT
NINLEKSKKKKKKKKK"
ORIGIN
Alignment Scores:
Pred. No.: 4.7e-100 Length: 1546
Score: 1116.00 Matches: 214
Percent Similarity: 92.56% Conservative: 10
Best Local Similarity: 88.43% Mismatches: 18
Query Match: 80.58% Indels: 0
DB: 10 Gaps: 0
US-10-649-273-2_COPY_148_414 (1-267) x BC078974 (1-1546)
QY 1 MetGluAlaHisAlaIleuThrIleArgIleuThrAsnIySGIlyPheProPheIleuVal 20
DB 585 ATGAGGCTCAAGCCCTACTATTTAGGCTGACCCCAAGTCGATTCCTTTTAGTT 644

```



QY 21 LeuLeuIIeSerGIyGIyhiScyLeuLeuAlaleuValGInGIyValIseRaspheLeu 40  
 DB 645 CTTTTCATTTCTGGAGGCCACTGCTGTTGGCGTTACTGACAGATTTTCTG 704  
 QY 41 LeuLeuGIyLysSerLeuAspIIeAlaProGIyAspMetLeuAspLysValAlaArg 60  
 DB 705 CTCCTGGAGGAGGCCCTGGACATAGCCGACGACATGCTTGCAAGGTTGGCAAGAGA 764  
 QY 61 LeuSerLeuIIeLysHisProGIyCysSerThreSerGIyGIyLysAlaIleGIuHis 80  
 DB 765 CTTTCTTAATCAACATCCAGATGTTCTACATGATGCTGGGAAGCTATGAACT 824  
 QY 81 LeuAlaLysGIyGIyLysArgPheHisPheAspIIeLysProProLeuHisHisAlaLys 100  
 DB 825 TTGGCCAAAGAAAGAAATAGATTCACACTTACTTACTATCAATCCACCATGCAAGATGTAAG 884  
 QY 101 AsnCyAspPheSerPheThrGIyLeuGIyHisValThrAspLysIleIleMetLysLys 120  
 DB 885 AACTGTGATTTTCTTTTACGGACCTTCACATGTCACCATTAAGCTAATACACACAG 944  
 QY 121 GluLysGIyGIyGIyLeuLysGIyGIyLeuSerSerAlaAlaAspIIeAlaAla 140  
 DB 945 GAAAAAGAAAGAGCATTTGAGAGGAGCAATCTGTATCATGCCGACCATTTGCT 1004  
 QY 141 ThrValGIyHisThrMetAlaCyHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
 DB 1005 GCGGTACAGACGCCACAGCGTGCACCTTGCGAAAAAACATCGTGTATTTCTGTT 1064  
 QY 161 CysLysGIyHisAspLeuLeuProGIyAsnAsnAlaValLeuValAlaSerGIyVal 180  
 DB 1065 TGCCACAGAAATTTGCTATCTCCAGCTACCGCATTAATGTTGTCTGGAGGTGTT 1124  
 QY 181 AlaSerAspPheThrIleArgArgAlaLeuGIyIleLeuThrAsnAlaThrGIyCysThr 200  
 DB 1125 GCAAGTAACTTGTACATCCAGAGCATGGAATTTGACAAATGGACCAATGCACT 1184  
 QY 201 LeuLeuCySerProProArgLeuCyThrAspAsnGIyIleMetIleAlaIleArgAsnGIy 220  
 DB 1185 TTGTTGTGTCCTCCCTCGAGACTGTGCATGCATGACATGATGATGATGATGAA 1244  
 QY 221 IleGIyArgLeuArgAlaGIyLeuGIyIleLeuHisAspIIeGIyIleArgTyGIy 240  
 DB 1245 ATGGAAGATTAAGTGTGCTGCTGGCATTTTACATGATGATGATGATGATGAA 1304  
 QY 241 ProLys 242  
 DB 1305 CCMAAG 1310  
 RESULT 14  
 BC051211 1109 bp mRNA linear ROD 15-APR-2003  
 LOCUS Mus musculus, clone IMAGE:1327545, mRNA.  
 DEFINITION BC051211  
 ACCESSION BC051211 GI:29881634  
 VERSION BC051211.1  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 Strausberg, R.  
 Direct Submission  
 Submitted (14-APR-2003) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: http://mgc.nci.nih.gov  
 Contact: MGC help desk  
 Email: cgabs-remail.nih.gov  
 Tissue Procurement: Marcello Bento Soares, Ph.D.  
 cDNA Library Preparation: M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILMIL)  
 DNA Sequencing by: Institute for Systems Biology

http://www.systemsbio.org  
 contact: amadan@systemsbio.org  
 Anup Madan, Jessica Fahey, Erin Helton, Mark Ketterman, Anuradha  
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/ILMIL at: http://image.llnl.gov  
 Series: IRAC Plate: 113 Row: b Column: 1  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Hexamer frequency ORF  
 analysis.  
 FEATURES  
 source location/Qualifiers  
 1..1109  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1327545"  
 /tissue\_type="Thymus gland, mouse"  
 /clone\_id="Soares\_thymus\_2nbwt"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 6,23e-84 Length: 1109  
 Score: 950.50 Matches: 190  
 Percent Similarity: 79.10% Conservative: 22  
 Best Local Similarity: 70.90% Mismatches: 35  
 Query Match: 68.63% Indels: 21  
 DB: Gaps: 4  
 US-10-649-273-2\_COPY\_148\_414 (1-267) x BC051211 (1-1109)  
 QY 18 PheLeuValLeuLeuIIeSer-----GIyGIyhiScyLeu----- 29  
 DB 82 TTTTTCATTTCTGGAGGCCACTGCTGTTGGCGTTACTGACAGATTTTCTG 141  
 QY 30 -----LeuAlaLeuValGInGIyValIseRaspheLeuLeuGIyLysSerLeuAsp 47  
 DB 142 TCTGTTGAAACGACTGCTCAATAGTGGGTTATCTTGTCTTAAT-----AAT 192  
 QY 48 IleAlaProGIyAspMetLeuAspLys-----ValAlaArg 59  
 DB 193 TTGCAGCCAAATATATTATTAAGAAAAAANGTAATGCTGTTTATATAGTGGCAAGA 252  
 QY 60 ArgLeuSerLeuIIeLysHisProGIyCysSerThreMetSerGIyGIyLysAlaIleGIu 79  
 DB 253 AGACTTCTTAATCAACATCCAGATGTTCTACATGATGATGATGATGATGAA 312  
 QY 80 HisLeuAlaLysGIyGIyLysArgPheHisPheAspIIeLysProProLeuHisHisAla 99  
 DB 313 CAGTTGGCCAAAGAGCAAGAAATGATTCATTTACTAATCAATCCACTATGAGATCT 372  
 QY 100 LysAsnCyAspPheSerPheThrGIyLeuGIyHisValThrAspLysIleIleMetLys 119  
 DB 373 AAGAAATGCAATTTTCTTTCACGGGACTTAACATATATGATGATGATGATGAA 432  
 QY 120 LysGIyLysGIyGIyGIyLeuLysGIyGIyLeuSerSerAlaAlaAspIIeAla 139  
 DB 433 AAGAAAAAGAAAGAGCATTTGAGAGGAGCAATTCGTATCATGCTGACAGACTTCT 492  
 QY 140 AlaThrValGInHisThrMetAlaCyHisLeuValLysArgThrHisArgAlaIleLeu 159  
 DB 493 GCTCGGTACGATGACAGCAAGCGCTGCCACTTGCAGAAACACATCGGCTATTTG 552  
 QY 160 PheCysLysGIyGIyAspLeuLeuProGIyAsnAsnAlaValLeuValAlaSerGIy 179  
 DB 553 TTTTGCAGAGAAATTTGCTTCTCCACCTTAAGCAGATTAATGTTATCTGAGCT 612  
 QY 180 ValAlaSerAspPheThrIleArgArgAlaLeuGIyIleLeuThrAsnAlaThrGIyCys 199  
 DB 613 GTTCAAGTAACTTGTACATCCGAAAGCATTTGGAATTTGCGAAATGCAAGCAGTGC 672

Qy 200 ThrLeuLeuCyseProProProArgLeuCyseThraAspanglylleMetlleAlaTyrAsn 219  
Db 673 ACCTGTGTGTGCACCTCCAGACAGCTGTGCATGACATGATGATGATGATGAT 732  
Qy 220 GATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 239  
Db 733 GGAATGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 792  
Qy 240 GATProLyseCyseProLeuGlyValAspIleSerLyseGlyValGlyAlaSerIleLys 259  
Db 793 GAACCAAAATGCTCTTGTGAGACATATCCAGAGAGTTGCAAGAGCTGCCATMAA 852  
Qy 260 ValProGlnLeuLyseMetGluile 267  
Db 853 GTACCCGATTAATAATGCACCT 876

RESULT 15  
BX934991 1522 bp mRNA linear VRT 02-FEB-2004  
LOCUS Gallus gallus finished cDNA, clone CHEST189114.  
DEFINITION BX934991  
ACCESSION BX934991 GI:41635519  
VERSION BX934991.1 GI:41635519  
KEYWORDS Gallus gallus (chicken)  
SOURCE Gallus gallus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 1522)

REFERENCE  
AUTHORS Boardman, P. B., Bonfield, J. K., Brown, W. R. A., Carder, C., Chalk, S. E.,  
Croming, M. D. R., Davies, R. M., Francis, M. D., Grafham, D. V.,  
Hubbard, S. J., Humphray, S. J., Hunt, P. J., Maddison, M., McLaren, S. R.,  
Noble, D., Overton, I. M., Rogers, J., Scott, C. E., Taylor, R. G.,  
Tickle, C. and Wilson, S. A.  
DIRECT SUBMISSION  
Submitted (02-FEB-2004) Sanger Institute, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: chickens@hms.unist.ac.uk  
COMMENT BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA  
sequencing project.  
This sequence is from the  
BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,  
from a library constructed by Elizabeth Bosch. cDNA was prepared  
from RNA extracted from muscle, normalised, and poly A-tailed.  
ECORI-NotI cut cDNA was then ligated into the vector. Vector:  
pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI Host: Escherichia  
coli DH10B.

TITLE  
JOURNAL  
COMMENT  
FEATURES  
source location/Qualifiers  
1.1522  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="layer and broiler"  
/db\_xref="taxon:9031"  
/clone="CHEST189114"  
/clone\_lib="CGEGRB11"  
/dev\_stage="adult"

ORIGIN  
Alignment Scores:  
Pred. No.: 1.24e-82 Length: 1522  
Score: 939.00 Matches: 176  
Percent Similarity: 81.06% Conservative: 38  
Best Local Similarity: 66.67% Mismatches: 50  
Query Match: 67.80% Indels: 0  
Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x BX934991 (1-1522)

Qy 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
Db 587 ATGAGGCTCAGCGACTTACATCAGACTGACAGAGCAAGTGAATTTCCCTTCTAGTT 646  
Qy 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40

Db 647 CTTTACTCTCCGAGGCTCAGCTGCACTTGGCGAGAGACAGAGGTTTCAGATTCTT 706  
Qy 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 60  
Db 707 CTGCTTGACAGCTCATGATATATACACAGAGTGCATGCTGGATTAAGTACAGAAAG 766  
Qy 61 LeuSerLeuIleLysHisProGlnCysSerThrMetSerGlyGlyLysAlaIleGluHis 80  
Db 767 CTCTCTTAGTGAAGCACCCGAGAGCCAGCGCATGCTGGGGGGAGGACATTAAGCAC 826  
Qy 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
Db 827 CTGGCTCAACCCGAGAGCTGCGAACAGTACACTTTCACACTTCCATGACAGTATGCT 886  
Qy 101 AsnCyseAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
Db 887 AACTGTGATTTTCTTTCTCCGACTTACAGCCCTTGTCACAAAGCCATTCTTCAGAA 946  
Qy 121 GluLysGlnGlyGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140  
Db 947 GAAAAAGAAAGAGTATTCAGAAAGGGGAAATCTGCTCGCTTAAGACATCCCTGCT 1006  
Qy 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgValIleLeuPhe 160  
Db 1007 GCTGCACAGCAGTATGCTGCTCATATATTCACCGGACACACCGCCATGCTCTTC 1066  
Qy 161 CysLysGlnArgAspLeuLeuProGlnHisAsnAlaValLeuValAlaSerGlyLysVal 180  
Db 1067 TGCATGAAACACGATATTTATACCAAAACCTGCACCTGTGGTATCAGAGAGATT 1126  
Qy 181 AlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200  
Db 1127 GCAAGTATACGATATATCAGAAAGAGCTGCAGACTGCGCAAGGCAAGCGTTTGGCT 1186  
Qy 201 LeuLeuCyseProProProArgLeuCyseThraAspanglylleMetlleAlaTyrAsn 220  
Db 1187 TTTCTGTCTCTCCCTCCAGAGCTGTGCACCGCATATGCTATGATTCATGAGATGCT 1246  
Qy 221 IleGluAlaGluArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGlu 240  
Db 1247 ATGAAAGATGCGCGAGAGTGTATTTATACAGTACTGAATGCGCATTCGACGA 1306  
Qy 241 ProLyseCyseProLeuGlyValAspIleSerLyseGlyValGlyAlaSerIleLysVal 260  
Db 1307 CCAAAAGCTCCCTTGAATGATATTTCCAAAAGAGTTGAAGAGGATTCATCAAGTG 1366  
Qy 261 ProGlnLeuLys 264  
Db 1367 CCAAGACTTAAG 1378

Search completed: June 16, 2005, 23:18:39  
Job time : 3202.26 secs



XX 30-JAN-2001; 2001US-00774528.  
 PR  
 XX (HYSE-) HYSECO INC.  
 PA (GOOD/) GOODRICH R. W.  
 PA  
 XX Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;  
 DR MPI; 2003-058563/05.

XX Novel polypeptide useful for treating neurodegenerative diseases, myeloid  
 PT or lymphoid cell disorders, bone disorders, mechanical and traumatic  
 PT disorders, coagulation disorders, and inflammatory diseases.  
 PS  
 PS Claim 1; Page: 612pp; English.

XX This invention relates to the cDNA sequences encoding an isolated novel  
 CC human polypeptide. The protein encoded by the nucleic acid of the  
 CC invention is useful for treating central and peripheral nervous system  
 CC diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic  
 CC lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,  
 CC Alzheimer's disease); autoimmune disease (e.g. systemic lupus  
 CC erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)  
 CC ; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopenia)  
 CC ; wounds, ulcers, burns; bone disorders (e.g. osteoporosis,  
 CC osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head  
 CC trauma); lung or liver fibrosis; reperfusion injury in various tissues;  
 CC bacterial, viral or fungal infections; allergic conditions such as  
 CC allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);  
 CC cancer and tumors; and inflammatory diseases (e.g. septic shock, Crohn's  
 CC disease, anaphylaxis). The protein may be used to inhibit the growth,  
 CC infection or function of infectious agents such as bacteria, fungi,  
 CC viruses, or to effect bodily characteristics, biorhythms or circadian  
 CC cycles of rhythms. The protein may also have  
 CC proliferation/differentiation, stem cell growth factor, haematopoiesis,  
 CC regulation, immune stimulation or suppressing, chemotactic/chemokinetic,  
 CC haemostatic and thrombolytic, receptor/ligand, and antiinflammatory  
 CC activities. The cDNA sequences of the invention are useful for expressing  
 CC recombinant protein for analysis. The present sequence represents a novel  
 CC human cDNA sequence of the invention, this sequence is an expressed  
 CC sequence tag (EST) and was identified using subtractive hybridization  
 XX  
 XX

SO Sequence 1416 BP; 441 A; 280 C; 273 G; 422 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 4.68e-147 Length: 1416  
 Score: 1385.00 Matches: 267  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x ABX70950 (1-1416)

QY 1 MetGluAlaHisAlaLeuThrIleAglLeuThAsnLysValGluPheProPheLeuVal 20  
 DB 502 ATGGAGGCTCATGCACTTACTATTAGGTTGACCAATGAAGAGAAATTTCTTTTGTGTT 561  
 QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValAspAspPheLeu 40  
 DB 562 CTTTTCATTTCTGGAGGCTCATGCTGTTGGCATTAAGGAGTTTTCAGATTTCCTG 621  
 QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
 DB 622 CTTCTTGGAAAGCTTTGGACATAGCACAGGTGACATGCTGACCAAGGTGGCAAGAAGA 681  
 QY 61 LeuSerLeuIleLysHisPProGluCysSerThrMetSerGlyGlyLysAlaIleGlnHis 80  
 DB 682 CTTTCTTTAATAAATCAATCAGAGAGTCTCCACCATGAGTGTGGCAAGCATAGAAAT 741  
 QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProPheLeuHisAlaLys 100

DB 742 TTGGCCAAACAGGAATAGATTTCATTGACATCAAACTCCCTTCATCATGCTTAA 801  
 QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysHis 120  
 DB 802 AATTGTGATTTTCTTTTACCTGACCTTCAACACTTCTGATTAATAATATGAAAAAG 861  
 QY 121 GluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140  
 DB 862 GAAAAGAGGAGATTTGAGAAAGGCAATCTGCTTCAGACAGACATGCTCTCC 921  
 QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
 DB 922 ACAGTACAGCACACAAAGGACATGCTGTGTGAAGAAACACATCGGGCTAATCTGTTT 981  
 QY 161 CysLysGlnArgAspIleLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180  
 DB 982 TGTAAAGAGAGAGCTGTGTATCTCAAAATATATGACATGCTGTGCTGTGCTGCTC 1041  
 QY 181 AlaSerAspPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200  
 DB 1042 GCAAGTAACTTCTATATCCGAGAGCTCTGGAATTTTAAACAAAGCAGACAGTCACT 1101  
 QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaITPAsnGly 220  
 DB 1102 TTGTTGTCTCTCTCCACAGACTATGACATGATATATGATATGATATGATGATGATGAT 1161  
 QY 221 IleGlnArgLeuAlaGlyIleGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgThrGlu 240  
 DB 1162 ATTAAAGACTACCTGCTGCTGGCAATTTTACATGACATAGAAAGGACATCCGCTATATA 1221  
 QY 241 ProLysCysProLeuGlyValAspIleSerLysGlnValGlyGlnAlaSerIleLysVal 260  
 DB 1222 CCAAAATCTCTCTTGGAGTATGACATATCAAAAGATGAGAGAGCTTCCATATAAAGTA 1281  
 QY 261 ProGlnLeuLysMetGluIle 267  
 DB 1282 CCACAAATTAATAAATGAGAGATA 1302

RESULT 2  
 ABS76639  
 ID ABS76639 standard; DNA; 1526 BP.  
 XX  
 AC ABS76639;  
 XX  
 DT 11-DEC-2002 (first entry)  
 XX  
 DE DNA encoding novel human metalloprotease MPI fragment #1.  
 XX  
 KW Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;  
 KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
 KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
 KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
 KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
 KW liver disease; renal disease; immune disorder; rheumatoid arthritis;  
 KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
 KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
 KW neurological disorder; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO200272751-A2.  
 XX  
 XX 19-SEP-2002.  
 PD  
 PF 05-FEB-2002; 2002WO-US003353.  
 XX  
 XX 05-FEB-2001; 2001US-0266518P.  
 PR 10-APR-2001; 2001US-0282814P.  
 XX  
 XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 PA  
 XX Chen J, Feder J, Nelson TC, Duclos F, Krystek S;  
 PI

XX WPI; 2002-723329/78.  
DR P-PSDB; ABG96487.  
XX  
XX New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
PT treating, or ameliorating diseases associated with aberrant  
PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and  
XX neurological disorders.  
XX  
XX Disclosure; Page 462-463; 473pp; English.  
XX  
XX The invention describes an isolated nucleic acid molecule (1) encoding a  
CC metalloproteinase (MP-1). (1) is useful for preventing, treating, or  
CC ameliorating a medical condition, particularly an immune disorder, an  
CC aberrant glutamate transport or motor neuron disorder, such as  
CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
CC condition. The compositions and methods are also useful for diagnosing,  
CC prognosticating, treating, ameliorating and/or treating disorders  
CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver,  
CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
CC Alzheimer's disease or Parkinson's disease). This sequence represents a  
CC metalloproteinase MPI polynucleotide  
XX  
XX Sequence 1526 BP; 484 A; 297 C; 289 G; 456 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 5,2e-147 Length: 1526  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservativity: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x ABS76639 (1-1526)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnValGluPheProPheLeuVal 20  
DB 1 ATGAGGCTCATGCACTTACTTATAGTTGACCAATTAAGTGAATTTCTTTTATGTT 60  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
DB 61 CTTTGTATTTCTGGAGTCACTGTCTGTGGCATTAGTTCAAGAGTTTCGATTTCCTG 120  
QY 41 LeuLeuGlyIleSerLeuAspPheIleAlaProGlyIleAspPheLeuValAlaArgArg 60  
DB 121 CTTCTTGAAAGCTCTTGGACATAGCACAGTGCATGCTTGCACAGGTGCAGAAAGA 180  
QY 61 LeuSerLeuIleIleHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 80  
DB 181 CTTTCTTAATAAACAATCCAGAGTCTCCACATGAGTGGGAAAGCCATGACAT 240  
QY 81 LeuAlaIleGlnGlyAsnArgPheHisPheAspIleIleValProLeuHisAlaIleVal 100  
DB 241 TTGGCCAAACAAGAAATAGATTTCATTTCATCAATCAAACTCCCTTCATCATGCTAAA 300  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspValIleIleMetLeuVal 120  
DB 301 AATTGTATTTTCTTTTACTGACCTTCAACACCTTACTGATTAATAATATATGAAAAAG 360  
QY 121 GluIleGlnGluGlyIleGluValGlyGlnIleLeuSerSerAlaAlaAspIleAlaIle 140  
DB 361 GAAAGAAAGAGGATATGAGAAAGGCAATCCGTCTTACGACGACGACATGCTCC 420  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValIleValArgThrHisArgAlaIleLeuPhe 160  
DB 421 ACACTACAGCACACAAATGCAATGATCTTTGTGAAAAAGAACATCGGCGTATTTCTGTT 480

QY 161 CysIleGlnArgAspLeuLeuProGlnAsnAsnAlaValIleLeuValAlaSerGlyVal 180  
DB 481 TGTAAAGCAGAGACTTGTACTTCAATAATATAGCACTGCTGCAATCTGCTGTC 540  
QY 181 AlaSerAsnPheIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200  
DB 541 GCAAGTACTTCTATATCCGAGACCTTGGAAATTTTAAACAAAGCAACAGTGCAT 600  
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220  
DB 601 TTGTTGTCTCTCCCTCCAGACTATGCACTGATATGTCATATGATTCATGCAATGCT 660  
QY 221 IleGluArgLeuAlaGlyIleLeuGlyIleLeuHisAspIleGluGlyIleArgGlyArg 240  
DB 661 ATTTGAAGACTACGCTGCTGGCATTTTACATGACATAGAAAGCATCGCTATGAA 720  
QY 241 ProIleCysProLeuGlyValAspIleSerIleGlyValGluAlaSerIleIleVal 260  
DB 721 CCAAAATCTCTTGGAGTGAACATATCAAAAGAGTTTGAGAAAGCTTCCATTAAGTA 780  
QY 261 ProGlnLeuIleMetGluIle 267  
DB 781 CCACATTAATAATGAGATTA 801  
RESULT 3  
ABA93268  
ID ABA93268 standard; cDNA; 2058 BP.  
AC ABA93268;  
XX  
XX 19-APR-2002 (first entry)  
XX  
XX Human O-sialoglycoproteinase-like protein encoding cDNA SEQ ID NO:1.  
XX  
XX Human, O-sialoglycoproteinase-like protein; OSGPLP; enzyme; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FT CDS 110..1354  
FT /tag= a  
FT /product= "O-sialoglycoproteinase-like protein"  
XX  
XX CN1318550-A.  
XX  
XX 24-OCT-2001.  
XX  
XX 19-APR-2000; 2000CN-00106834.  
XX  
XX 19-APR-2000; 2000CN-00106834.  
XX  
XX (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.  
XX  
XX Mao Y, Xie Y;  
XX  
XX WPI; 2002-115090/16.  
XX  
XX P-PSDB; ABB05481.  
XX  
XX O-sialoglycoproteinase-like protein and encoding polynucleotide, useful  
PT for diagnosing, preventing and treating related diseases.  
XX  
XX  
XX Claim 5; Page 29-30 (Disclosure); 38pp; Chinese.  
XX  
XX The present sequence encodes human O-sialoglycoproteinase-like protein  
CC (OSGPLP). The present invention also describes: (1) the preparation of  
CC the OSGPLP protein; (2) applying the OSGPLP protein in diagnosis; (3) the  
CC prevention and/or treatment of related diseases; (4) utilizing the OSGPLP  
CC protein in screening its agonist, excitomoter and inhibitor and preparing  
CC an antibody against the OSGPLP protein; and (5) the use of the OSGPLP  
CC polynucleotide sequences, proteins, agonists, excitomoters, inhibitors  
CC and antibodies in treating diseases related to the abnormal OSGPLP gene  
CC and in preparing the medicine composite for the treatment  
XX

Sequence 2058 BP, 637 A, 400 C, 410 G, 611 T, 0 U, 0 Other;

## Alignment Scores:

|                        |           |               |      |
|------------------------|-----------|---------------|------|
| Pred. No.:             | 7,97e-147 | Length:       | 2058 |
| Score:                 | 1385.00   | Matches:      | 267  |
| Percent Similarity:    | 100.00%   | Conservative: | 0    |
| Best Local Similarity: | 100.00%   | Mismatches:   | 0    |
| Query Match:           | 100.00%   | Indels:       | 0    |
| DB:                    | 6         | Gaps:         | 0    |

US-10-649-273-2\_COPY\_148\_414 (1-267) x ABA93268 (1-2058)

```
OY 1 MetGluaIahisAlaLeuThriIeaArgLeuThraAsnlyValGluPheProPheLeuVal 20
DB 551 ATGGAGGCTCATGCTACTTACTTATAGTTGACCAATTAAGATTAAGATTCTTTTATGTT 610
OY 21 LeuLeuIIsSerGlyGlyYhisCysLeuLeuAlaLeuValGlnGlyValISerAspPheLeu 40
DB 611 CTTTGTATTTCTGAGGTCACCTGCTGTTGGCAATTAGTTCAGAGATTTCAGATTTCCTG 670
OY 41 LeuLeuGlyIlySerLeuAspIleAlaProGlyAspMetLeuAspIlyValIAspArg 60
DB 671 CTTCTTGGAAGCTTTTGACATAGCACAGGTGACATGCTTGACAAAGTGGCAAGAGA 720
OY 61 LeuSerLeuIlelyshISProGlyCysSerThMetSerGlyGlyValaIleGluHis 80
DB 731 CTTTCTTAATAAATCAATCCAGAGTCTCCACCATAGTGTGGGAAGCCATAGAACAT 790
OY 81 LeuAlaIySGlnGlyAsnAlyPheHisIspheAspIleIlyProPheLeuHisIAspAla 100
DB 791 TTGGCCAAACAAGGAATAGATTTCATTTGACATCAAACTCCCTTGACATGCTAAA 850
OY 101 AsnCyAspPheSerPheThrglyLeuGlnHisValThraSplyIleIleMetIlylys 120
DB 851 AATTGATTTTCTTTTACTGACCTTCAACAGCTTATCAATAAATAAATGAAGAAAG 910
OY 121 GluIySGlnGlyIleGluIySGlnIleLeuSerSerAlaIAspIleAlaIa 140
DB 911 GAAAAAGAGAGATTTGAGAAAGGGCAATCTGCTTCAGCAGACGATTTGCTGCC 970
OY 141 ThrValGlnHisThrMetAlaCysHisLeuValIlySargThraIleLeuPhe 160
DB 971 ACAGTACAGACACAAATGCGATGCTCTGTGAAAGAACACATCGGCTATTCTGTT 1030
OY 161 CysIySGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
DB 1031 TGTAAACAGAGACACTGTTACTCTCAAAATATGCACTGCTGTCATGTGGTGGTGC 1090
OY 181 AlaSerAspPheThryIleArgArgAlaLeuGlnIleLeuThraSnaIaThrgInCysThr 200
DB 1091 GCAAGTAACTTCTTAATCCGAGAGCTCTGAAATTTTAAACAAACGACACAGTGCAT 1150
OY 201 LeuLeuCySProProProArgLeuCySThrAspAsnGlyIleMetIleAlaIAspAsnGly 220
DB 1151 TTGTTGTGCTCTCTCCACACATATGACATGATATGACATTATGATGATGAAGGT 1210
OY 221 IIEGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyGln 240
DB 1211 ATTTGAAAGACTAGCTGCTGCTGGCATTTTACATACATAGACAGCATCCGCTATGAA 1270
OY 241 ProIyCySProLeuGlyValaAspIleSerIyGlnValaGlyIleAspIleIyVal 260
DB 1271 CCAAAATGCTCTTGGAGTAGACATATCAAAAGAAAGTGGAGAGCTTCATATAAAGTA 1330
OY 261 ProGlnLeuIyMetGluIle 267
DB 1331 CCAACATTAATAATGAGATTA 1351
```

RESULT 4  
ABST76635 standard; DNA; 2197 BP.  
XX  
AC ABST76635;

XX 11-DEC-2002 (first entry)  
DT  
XX DNA encoding novel human metalloprotease MP1.  
DE

KM Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;  
KM motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
KM reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
KM genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
KM Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
KM Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
KM liver disease; renal disease; immune disorder; rheumatoid arthritis;  
KM acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
KM emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
KM neurological disorder; gene; ds.

XX Homo sapiens.

XX W0200272751-A2.

XX 19-SEP-2002.

XX 05-FEB-2002; 2002W0-US003353.

XX 05-FEB-2001; 2001US-0266518P.

XX 10-APR-2001; 2001US-0282814P.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

XX Chen J, Feder J, Nelson TC, Duclos F, Krystek S;

XX WPI: 2002-723329/78.

XX P-PSDB: ABG96478.

XX New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
PT treating, or ameliorating diseases associated with aberrant  
PT metalloprotease activity, e.g. immune, metabolic, inflammatory and  
PT neurological disorders.

XX Claim 1; Fig 1A-C; 473bp; English.

XX The invention describes an isolated nucleic acid molecule (i) encoding a  
CC metalloprotease (MP-1) (i) is useful for preventing, treating, or  
CC ameliorating a medical condition, particularly an immune disorder, an  
CC aberrant glutamate transport or motor neuron disorder, such as  
CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
CC condition. The compositions and methods are also useful for diagnosing,  
CC prognosticating, treating, ameliorating and/or treating disorders  
CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
CC Alzheimer's disease or Parkinson's disease). This sequence represents a  
CC metalloprotease MP1 polynucleotide

XX Sequence 2197 BP, 681 A, 441 C, 439 G, 636 T, 0 U, 0 Other;

## Alignment Scores:

|                        |           |               |      |
|------------------------|-----------|---------------|------|
| Pred. No.:             | 8,75e-147 | Length:       | 2197 |
| Score:                 | 1385.00   | Matches:      | 267  |
| Percent Similarity:    | 100.00%   | Conservative: | 0    |
| Best Local Similarity: | 100.00%   | Mismatches:   | 0    |
| Query Match:           | 100.00%   | Indels:       | 0    |
| DB:                    | 6         | Gaps:         | 0    |

US-10-649-273-2\_COPY\_148\_414 (1-267) x ABST76635 (1-2197)

```
OY 1 MetGluaIahisAlaLeuThriIeaArgLeuThraAsnlyValGluPheProPheLeuVal 20
DB 551 ATGGAGGCTCATGCTACTTACTTATAGTTGACCAATTAAGATTAAGATTCTTTTATGTT 610
```

DB 672 ATGGAGGCTTCAGTACTTACTATTAGGTGACCAATAAGTAGAATTTCTTTTAGTT 731  
QY 21 LeuLeu11eSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValIleSerAspPheLeu 40  
DB 732 CTTTATATTTCTGAGAGTCACTGCTCTGTTGGCAATTACTTACAGAGATTTCGATTTTCTG 791  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaPheProGlyAspMetLeuAspLysValAlaArgArg 60  
DB 792 CTTCTTGGAAAGCTTTGACATAGCACAGGTGACATGCTTGACAAAGTGGCAAGAAGA 851  
QY 61 LeuSerLeu11eLysHisPheProGlyCysSerThrMetSerGlyGlyValAlaIleGlnHis 80  
DB 852 CTTTCTTTATATAAACAATCCAGAGTGTCTCCACCAATGATGTTGGAAAGCCATGAACAT 911  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysPheProLysHisAlaLys 100  
DB 912 TTGGCCAAAGCAAGATTAATTAATTCATTTTGACATCAAACTCCCTTGACATGCTTAA 971  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
DB 972 AATTGATATTTCTTTTACTGACCTTCAACAGCTTCTGATAAATATATATGAAAAAG 1031  
QY 121 GlnLysGlnGlyGlyIleGlyLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140  
DB 1032 GAAAAAGAGAGATTTGAGAAAGGCAAAATCTGCTTCAGACAGCAGCATTTGCTGCC 1091  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisSerAlaIleLeuPhe 160  
DB 1092 ACGGTACAGACACAAATGAGATGATCTGTGTGAAAGAACACATCGGGCTATTCCTGTT 1151  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 180  
DB 1152 TGTAAGCAGAGAGCTTTGTTACTCCTCAAAATATATCACTAGCTGCTGCTGCTGCTC 1211  
QY 181 AlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200  
DB 1212 GCAAGTACTCTTATATATCCGACAGCTCTGGAAATTTTAAACAAACGACACAGTCACT 1271  
QY 201 LeuLeuCysProProPheArgLeuCysThrAspAsnGlyIleMetIleAlaTTPAsnGly 220  
DB 1272 TTGTTGTGCTCTCTCCACAGCTATGACATGATATGATGATGATGATGATGATGAT 1331  
QY 221 IleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGln 240  
DB 1332 ATTGAAGACTACCTGCTGGCTGGCAATTTTACATGATCATTAGAAAGGATCCGCTATGAA 1391  
QY 241 ProLysCysProLeuGlyValAlaSerLysGlyValGlyGlnAlaSerIleLysVal 260  
DB 1392 CCAAAATGCTCTCTTGGAGTATGACATATCAAAAGAGTTGAGAGAGCTTCCATATAAGTA 1451  
QY 261 ProGlnLeuLysMetGlnIle 267  
DB 1452 CCACAATTAATAAATGAGAGAT 1472

RESULT 5  
ABT23207  
ID ABT23207 standard; DNA, 2572 BP.  
XX  
XX ABT23207;  
XX  
XX 01-MAY-2003 (first entry)  
DE Human protein modification + maintenance molecule DNA SEQ ID No 36.  
XX  
XX Cytosolic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;  
KW cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;  
KW antitumor; hepatotropic; gynecological; antibacterial; virucide;  
KW protozoacide; antiparasitic; cell proliferative disease; PMOD;  
KW protein modification and maintenance molecule; immunogenic fragment;  
KW cancer; autoimmune; inflammatory disease; neurological disorder;  
KW gastrointestinal; developmental; vesicle trafficking disorder; infection;  
KW protein-protein interaction; drug-target interaction;  
KW gene expression profile; human; gene; ds.

XX XX Homo sapiens.  
OS  
XX MO2003000844-A2.  
PN  
XX 03-JAN-2003.  
PD  
XX 18-JUN-2002; 2002MO-US019360.  
PF  
XX 22-JUN-2001; 2001US-0300508P.  
PR 06-JUL-2001; 2001US-0303445P.  
PR 13-JUL-2001; 2001US-0305405P.  
PR 09-AUG-2001; 2001US-0311442P.  
PR 24-AUG-2001; 2001US-0314821P.  
PR 29-AUG-2001; 2001US-0315992P.  
PR 03-MAY-2002; 2002US-0378205P.  
XX (INCY- ) INCYTE GENOMICS INC.  
PA  
XX Gandhi AR, Kable AE, Swarnakar A, Hafalia AUA, Tran B, Duggan BM,  
PI Warren BA, Ison CH, Honchell CD, Nguyen DB, Lu DM, Lee EA, Yue H,  
PI Forsythe TJ, Barroso I, Ramkumar J, Griffin JA, Li JX, Yang J,  
PI Thangavelu K, Gietzen KJ, Ding L, Baughn MR, Borowsky ML, Yao MG,  
PI Walla NK, Mason PM, Gurnurajan R, Lee S, Becha SD, Lee ST, Tran UK,  
PI Elliott VG, Luo W, Sprague WW, Tang YT, Lu Y, Zebaryadian Y,  
XX WPI: 2003-184039/18.  
DR P-FSDB; ABJ26654.  
XX  
XX New isolated human PMOD polypeptide and polynucleotide, useful for  
PT diagnosing, treating and preventing diseases or conditions associated  
PT with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and  
PT infections.  
XX  
XX Claim 91; Page 211; 225pp; English.  
PS  
XX The invention relates to an isolated polypeptide comprising: any of 28  
CC sequences of 48-1256 amino acids; a natural amino acid sequence at least  
CC 90% identical to the 28 amino acid sequences, 94% identical to a sequence  
CC of 703 or 267 amino acids, 96% identical to a sequence of 414 amino  
CC acids, or 97% identical to a sequence of 242 amino acids, all given in  
CC the specification; or a biologically active or immunogenic fragment of  
CC the isolated polypeptide. The polypeptides and polynucleotides are useful  
CC in diagnosing, treating and preventing diseases or conditions associated  
CC with the decreased expression of protein modification and maintenance  
CC molecules (PMOD), such as cell proliferative diseases (e.g. cancer,  
CC atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,  
CC allergies), neurological disorders (e.g. stroke, Parkinson's disease,  
CC epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.  
CC endometriosis), developmental, vesicle trafficking disorders, and  
CC infections (e.g. bacterial, viral, parasitic, protozoal). These are also  
CC useful in assessing the effects of exogenous compounds on the expression  
CC of nucleic acid and amino acid sequences of PMOD. The PMOD or its  
CC fragments are useful in screening compounds for effectiveness as agonist  
CC or antagonist of the polypeptides, or in altering the expression of the  
CC target polynucleotide and compounds that specifically bind to or modulate  
CC the activity of the polypeptide. The microarray is useful in monitoring  
CC or measuring protein-protein interactions, drug-target interactions, and  
CC gene expression profiles. This polynucleotide sequence represents the DNA  
CC encoding a human PMOD protein of the invention  
SQ Sequence 2572 BP; 780 A; 489 C; 509 G; 794 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 1,09e-146 Length: 2572  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 3  
DB: 8 Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x ABT23207 (1-2572)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
 DB 585 ATGGAGGCTCATGCCTTACTATTAGGTGACCAATTAAGTAGAATTTCTTTTGTAGTT 644  
 QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAPheLeu 40  
 DB 645 CTTTGTATTTCTGGAGTCACTGTCTGTGGCAATTGTTCAAGAGATTTCAGATTTCGTG 704  
 QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
 DB 705 CTTCTGGAAAGCTTTGGACATACGACCGAGTGACATGCTTACACAGGTGGCAAGAGA 764  
 QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyLysAlaIleGluHis 80  
 DB 765 CTTTCTTATATAAACATCATCAGAGTCTCCACCATAGTGTGGAAAGCCATAGAACAT 824  
 QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
 DB 825 TTGGCCCAACCAAGCAATAGATTTCATTGTGACATCAACCTCCCTTGACATCATGTCTAAA 884  
 QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
 DB 885 AATTGTGATTTTCTTTTACTGACATTCACACGTTACTGATTAATTAATTAATGAAG 944  
 QY 121 GlyLysGluGlnGlyIleGlyLysGlnIleLeuSerSerAlaAlaAspIleAlaAla 140  
 DB 945 GAAAAAGAGAGAGATTGAGAAAGGCAAAATCTGTCTTCACACAGACATTTGCTGCC 1004  
 QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
 DB 1005 ACAGTACACACACCAATGCGATCTCATCTGTGAAAAAACAATCGCGCTATTCTGTTT 1064  
 QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 180  
 DB 1065 TGTAAAGCAAGAGACTGTGCTCAAAATTAATGACGACTGTGCTGTGCTGTGCTGTC 1124  
 QY 181 AlaSerAspPheTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200  
 DB 1125 GCAAGTAATCTCATATCCGACAGAGCTCGAATTTTAAACAAACGACACAGTGCAT 1184  
 QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaIlePheGly 220  
 DB 1185 TTGTTGTGCTCTCTCCACAGCATATGACATGATTAATGCAATTAATGATGATGAATGCT 1244  
 QY 221 IleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyGln 240  
 DB 1245 ATTTGAAAGACTAGCTGTCTGTGGCATTTTACATACATTAAGAGCATCCGCTAAGAA 1304  
 QY 241 ProLysCysProLeuGlyValAlaAspIleSerLysGlnValGlyGluAlaSerIleLysVal 260  
 DB 1305 CCAAAATGCTCTTGTGAGTAGACATATCAAAAGAAAGTTGAGAGAGCTTCATTAAGTA 1364  
 QY 261 ProGlnLeuLysMetGluIle 267  
 DB 1365 CCACATTAATAAATGAGATA 1385  
 DB  
 RESULT 6  
 AAD46856  
 ID AAD46856 standard; cDNA; 1820 BP.  
 AC AAD46856;  
 XX  
 DT 27-JAN-2003 (first entry)  
 XX  
 XX Human glycoprotease 28472 cDNA.  
 DE  
 XX Human; adenosine deaminase; seven transmembrane domain receptor; cancer;  
 KW 7TM; glycoprotease; immune disorder; iga deficiency; allergy; arrhythmia;  
 KW rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular disorder;  
 KW hypertension; ischemic heart disease; obesity; myocardial infarction;  
 KW endothelial cell disorder; Grave's disease; psoriasis; brain disorder;  
 KW Parkinson's disease; Alzheimer's disease; haematopoietic disorder;  
 KW cerebral oedema; metabolic disorder; liver disorder; platelet disorder;

KW chromosome mapping; tissue typing; gene therapy; neuroprotective;  
 KW cytoskeletal; anorectic; cardiant; haemostatic; gene; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 146..1390  
 FT /\*tag= a  
 FT /product= "Human 28472 protein"  
 FT /note= "This region is specifically claimed as SEQ ID NO:  
 6 in claim 1 of the specification"  
 PN WO200274960-A2.  
 PD 26-SEP-2002.  
 XX  
 XX 08-NOV-2001; 2001WO-US051427.  
 PF  
 PF 08-NOV-2000; 2000US-0246768P.  
 PR 08-NOV-2000; 2000US-0246772P.  
 PR 15-NOV-2000; 2000US-0249185P.  
 XX  
 PA (MILL.) MILLENNIUM PHARM INC.  
 PI  
 PI Leiby KR, Kapeller-Libermann R, Glucksmann M;  
 XX  
 XX WPI: 2002-759898/82.  
 DR P-PSDB; AAE29234.  
 DR  
 XX  
 XX New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,  
 PT useful for diagnosing and treating cancer, immune, cardiovascular,  
 PT hematopoietic, brain, pain, metabolic, liver or platelet disorders, and  
 PT in pharmacogenomics.  
 PS Claim 1, Fig 8, 178pp; English.  
 XX  
 XX The present invention relates to novel 38650, 28472, 5495, 65507, 81588  
 CC or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-  
 CC protease or seven transmembrane domain (7TM) receptor family members.  
 CC Sequences of the invention are useful in diagnosing and treating cancer  
 CC or aberrant cellular proliferation and/or differentiation (e.g. colon or  
 CC lung cancer), immune disorders (e.g. selective iga deficiency, rheumatoid  
 CC arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.,  
 CC hypertension, atherosclerosis, arrhythmias, ischemic heart disease,  
 CC myocardial infarction, thrombus) including endothelial cell disorders  
 CC (e.g. psoriasis, Grave's disease), haematopoietic disorders, brain  
 CC disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease),  
 CC pain and metabolic disorders (e.g. obesity), liver disorders or platelet  
 CC disorders. They are also useful in screening assays, predictive medicine  
 CC (e.g. diagnostic assays, prognostic assays, monitoring clinical trials  
 CC and pharmacogenetics) and prophylactic and therapeutic methods. The  
 CC nucleic acids may also be used in chromosome mapping, tissue typing and  
 CC forensic biology and as surrogate markers. Sequences of the invention are  
 CC also used in gene therapy. The present sequence is human glycoprotease  
 CC 28472 cDNA  
 XX  
 SQ Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;  
 Alignment Scores:  
 Pred. No.: 1,07e-142 Length: 1820  
 Score: 1348.00 Matches: 260  
 Percent Similarity: 98.50% Conservative: 3  
 Best Local Similarity: 97.38% Mismatches: 4  
 Query Match: 97.33% Indels: 0  
 DB: Gaps: 0  
 US-10-649-273-2\_COPY\_148\_414 (1-267) x AAD46856 (1-1820)  
 QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
 DB 587 ATGGAGGCTCATGCCTTACTATTAGGTGACCAATTAAGTAGAATTTCTTTTGTAGTT 646  
 QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAPheLeu 40



```
DB 647 CTTTGGATTCTGAGAGTCACTGCTGTTGGCACTTACATTCAGAGATTTCCTG 706
QY LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspValAlaArg 60
DB 707 CTTCTTGAAGAGCTTTGGACATGACACAGGTGACATGCTTGACAGGTGGCAAGAGA 766
QY LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyValAlaGluHis 80
DB 767 CTTTCTTATATAAATCATCCAGAGTGTCTCCACCATGATGCTGGGAAGCATAGAACAT 826
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProPoleuHisHisAlaLys 100
DB 827 TTGGCCAAACAAGGAATAGATTCAATTTGACATCAAACTCCCTTGACATCAATGCTAAA 886
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
DB 887 AATTGTGATTTTCTTTTACTGACCTTCAACAGCTTACTATATAAATAATGAACACAG 946
QY 121 GluLysGlnGlyLysGlnGlyLysGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
DB 947 AAACAAGAGAGAGCTATGGAAGGGCAATCTGCTTCAGACAGACATGCTGCC 1006
QY 141 ThrValGlnHisThrMetAlaCysHisLysValLysArgThrHisArgAlaIleLeuPhe 160
DB 1007 AAGTACAGACACAATAGCATGATCTTGTGAAAAGAACACATCGGCTATTCTGTTT 1066
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
DB 1067 TGTAAACAGAGACACTTGTATCTCCAAATAATGCACTGCTGCTGCTGCTGCTGCTGCT 1126
QY 181 AlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
DB 1127 GCAAGTAACTCTTATATCCGACAGCTCTGAAATTTTAAACAAACGACACAGTGAAC 1186
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220
DB 1187 TTGTTGTGCTCTCTCCACAGCATATGACATGATATGACATTAATGATGATGAATGCT 1246
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTrpGlu 240
DB 1247 ATGAAAGACTACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1306
QY 241 ProLysCysProLeuGlyValAspIleSerLysGlnValGlyGlnAlaSerIleLysVal 260
DB 1307 CCAAAATGCTCTTGGAGTGAACATATCAAAAAGAGTGGAGAAGCTTCCATTAAGAATA 1366
QY 261 ProGlnLeuLysMetGluIle 267
DB 1367 CCACATTTAAAAATGAGAGATA 1387
RESULT 7
ACA60887
ID ACA60887 standard; cDNA, 1820 BP.
AC ACACA60887;
AC ACACA60887;
DE 08-JUL-2003 (first entry)
XX Human cDNA 28472 encoding a glycoprotease.
XX
XX
XX Human; ss; gene; cancer; aberrant cellular proliferation;
XX differentiaction; immune disorders; heart disorder; brain disorder;
XX cardiovascular disorder; endothelial cell disorder; pain disorder;
XX haematopoietic disorder; blood vessel disorder; metabolic disorder;
XX liver disorder; platelet disorder; glycoprotease.
XX
XX Homo sapiens.
XX
XX KEY location/Qualifiers
XX CDS 146..1390
XX /tag= a
XX /product= "Glycoprotease"
```

```
FT /note= "this CDS is specifically claimed in claim 1"
XX
XX US2003009017-A1.
XX
XX 09-JAN-2003.
XX
XX 08-NOV-2001; 2001US-00012140.
XX
XX 08-NOV-2000; 2000US-0246768P.
XX 08-NOV-2000; 2000US-0246772P.
XX 15-NOV-2000; 2000US-0249185P.
XX
XX (LEIBY/ LEIBY K R.
XX (KAPE/ KAPPELLER-LIBERMANN R.
XX (GLUC/ GLUCKSMANN M A.
XX
XX Leiby KR, Kapeller-Libermann R, Glucksmann MA;
XX WPI; 2003-428888/40.
XX P-PSDB; ABU09569.
XX
XX New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid
XX molecules, useful for diagnosing, treating cancer, pain, or immune
XX heart, endothelial cell, hematopoietic, blood vessel, brain, metabolic
XX and liver disorders.
XX
XX Claim 2; Fig 8; 90pp; English.
XX
XX The invention relates to an isolated 38650 (encoding adenosine
XX deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7
XX transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or
XX a sequence which is at least 60% identical to the six nucleic acids or
XX their open reading frames, fragments of at least 15 nucleotides,
XX naturally occurring variants, or a DNA insert of the plasmid deposited
XX with the American Type Culture Collection as Accession No. not defined in
XX the specification, which encodes the amino acid sequence). Also included
XX are a host cell containing the nucleic acids (used to produce the
XX proteins), the encoded proteins, an antibody that selectively binds to
XX the polypeptide, and identifying a compound that binds to/modulates the
XX activity of the polypeptide. The nucleic acid molecules, polypeptides and
XX methods are useful for diagnosing, treating cancer, aberrant cellular
XX proliferation and/or differentiation, immune disorders, heart disorders,
XX cardiovascular disorders including endothelial cell disorders,
XX haematopoietic disorders, blood vessel disorders, brain disorders, pain
XX and metabolic disorders, liver disorders and platelet disorders (many
XX examples of these disorders are given in the specification). The present
XX sequence is the Human cDNA 28472 encoding a glycoprotease
XX
XX SQ Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;
XX
XX Alignment Scores:
XX Pred. No.: 1,07e-142 Length: 1820
XX Score: 1348.00 Matches: 260
XX Percent Similarity: 98.50% Conservative: 3
XX Best Local Similarity: 97.38% Mismatches: 4
XX Query Match: 97.33% Indels: 0
XX DB: 8 Gaps: 0
XX
XX US-10-649-273-2_COPY_148_414 (1-267) x ACA60887 (1-1820)
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGlnPheProPheLeuVal 20
DB 587 ATGAGAGCTCATGACCTTACTATTAGGTGACCAATTAAGTAGAATTTCTTTTATAGT 646
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 647 CTTTGAATTTCTGAGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 706
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspValAlaArg 60
DB 707 CTTCTTGAAGAGCTTTGGACATGACACAGGTGACATGCTTGACAGGTGGCAAGAGA 766
QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyValAlaGluHis 80
```

```
DB 767 CTTCTTTAATAAATCAAGTCTCCACATGATGTCGAGAAAGCATAGAACAT 826
QY 81 LeuAlaIySGInGIyAsnARgPhehISpneASpIIeIySProProluEHIShISAlaIyS 100
DB 827 TTGGCCAAACAAAGGAATAGATTTCATTTTGACATCAAACTCCCTTGACATGCTAA 886
QY 101 AasnCyASpPheSerPheThrgIyLeuGInHIShISValIthraSpIySIIeIImetLys 120
DB 887 AATTGGATTTTCTTTTACTGATTCACACGTTACTGATTAATAATATGAAAACAG 946
QY 121 GIluySGInGIyIleGIuIySGInIleIeSerSerIaIaAspIIeIaIa 140
DB 947 AAACAGAGAGAAAGTATGAGAAAGGAGCAATCTGCTTCACACAGACATTCGCTGCC 1006
QY 141 ThrValGInHIShISrMetAlaCySHISleValIySaRgThrHISaRgAlaIleIePhe 160
DB 1007 AAGNTRACAGACACATGCGATCTCATCTTTGTAAGAAAGAACACATCGGCTATTCGTTT 1066
QY 161 CysIySGInARgAspleuEupProGInAsnAsnAlaValIeValAlaIaSerGIyGIyVal 180
DB 1067 TGTAAAGCAGAGACTGTTTACTCTCAAAATATATGACGATCTGGTTCATCGGAGGTC 1126
QY 181 AlaSerAspPheTyrIleARgATgAlaIeUgInIleUthrAsnAlaThrgInCySThr 200
DB 1127 GGAAGTAACTTCATATTCGACAGAGCTCTCGAAATTTTAAACAAACGACACAGTGCAT 1186
QY 201 LeuLeuCySProProARgLeuCySThrAspAsnGIyIleMetIleAlaTPAsnGIy 220
DB 1187 TTGTTGTGTCCTCTCCAGACTATGACATGATTAATGAGCATATGATGATGAAAGT 1246
QY 221 IlegIuARgLeuARgAlaGIyLeuGIyIleLeuHISaSpIIeGIyIleARgYrGIu 240
DB 1247 ATTGAAAGCTAGTCTGCTGGCTTTTACATGACATAGAACGATCCGCTATGAA 1306
QY 241 ProIyCySProLeuGIyValAspIIeSerIySGIuValGIyAlaSerIIeIyVal 260
DB 1307 CCAAAATGCTCTTTCGAGTACATATCAAAAGAGTGGAGAGCTTCATATAAAGTA 1366
QY 261 ProGInLeuIyMetGIuIle 267
DB 1367 CCACATTAATAAATGAGATA 1387
RESULT 8
ID ABS57020 standard; cDNA; 1821 BP.
AC ABS57020;
XX 30-JAN-2003 (first entry)
DE cDNA encoding novel human glycoprotease 28472.
XX
KW Cancer; aberrant cell proliferation; aberrant cell differentiation;
KW breast cancer; ovarian cancer; prostate cancer; colon cancer;
KW lung cancer; immune disorder; heart disorder; cardiovascular disorder;
KW endothelial disorder; hematopoietic disorder; blood vessel disorder;
KW brain disorder; pain; metabolic disorder; liver disorder; diabetes;
KW platelet disorder; carcinoma; sarcoma; leukemia; Hodgkin's disease;
KW autoimmune disorder; hypertension; atherosclerosis; heart failure;
KW myocardial infarction; ischemic heart disease; Crohn's disease;
KW Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;
KW cerebral ischemia; peripheral neuropathy; Alzheimer's disease;
KW Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; gene; ss.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 147..1391
FT /tag= a
FT /product= "glycoprotease 28472"
FT /note= "Specifically claimed in claim 1"
```

```
PN K0200277233-A2.
XX
PD 03-OCT-2002.
XX
PF 08-NOV-2001; 2001WO-US046724.
XX
PR 08-NOV-2000; 2000US-0246768P.
PR 08-NOV-2000; 2000US-0246772P.
PR 15-NOV-2000; 2000US-0249185P.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Leiby KR, Kapeller-Libermann R, Glucksmann M;
XX
DR WPI; 2003-029938/02.
DR P-PSDB; ABG71162.
XX
PT New adenosine deaminase, glycoprotease and seven transmembrane domain
PT nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,
PT 81588 and 14354, useful for treating e.g. leukemias, Hodgkin's disease or
PT hypertension.
XX
PS Claim 2; Fig 8A-B; 178pp; English.
XX
CC The invention describes isolated 38650, 28472, 5495, 65507, 81588 and
CC 14354 nucleic acid molecules (I) and their encoded polypeptides (II). The
CC 38650 nucleic acid molecule comprises a sequence encoding adenosine
CC deaminase. The 28472 nucleic acid molecule comprises a sequence encoding
CC a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise
CC sequences that encode a human seven transmembrane domain (7TM). The
CC 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide
CC sequences are useful for diagnosing, preventing or treating a subject
CC with or at risk of developing a disorder, e.g. cancer or aberrant
CC cellular proliferation and/or differentiation (e.g. breast, ovarian,
CC prostate, colon or lung cancer), immune disorders, heart disorders,
CC cardiovascular disorders, endothelial disorders, hematopoietic disorders,
CC blood vessel disorders, brain disorders, pain and metabolic disorders,
CC liver disorders or platelet disorders. These disorders include carcinoma,
CC sarcoma, leukemia, Hodgkin's disease, autoimmune disorders,
CC hypertension, atherosclerosis, heart failure, myocardial infarction,
CC ischaemic heart disease, Crohn's disease, Grave's disease, Kawasaki
CC syndrome, Raynaud's disease, aneurysm, cerebral ischemia, peripheral
CC neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,
CC cachexia or diabetes. This sequence encodes the novel human glycoprotease
CC 28472
XX
SQ Sequence 1821 BP; 543 A; 365 C; 394 G; 518 T; 0 U; 1 Other;
XX
Alignment Scores:
Pred. No.: 1.07e-142 Length: 1821
Score: 1348.00 Matches: 260
Percent Similarity: 98.50% Conservative: 3
Best Local Similarity: 97.38% Mismatches: 4
Query Match: 97.33% Indels: 0
DB: Gaps: 0
US-10-649-273-2_COPY_148_414 (1-267) x ABS57020 (1-1821)
QY 1 MetGInAlaHISaIeUthrIleARgLeuThraSnyValGIuPheProPheUVal 20
DB 588 ATGAGGCTCATGCACTTACTATTAGGTTACCAATTAAGTAGAATTTCTTTTATGTT 647
QY 21 LeuLeuIleSerGIyGIyHISCySleuLeuAlaIeUValGInGIyValISerAspPheU 40
DB 648 CTTTGAATTTCTGAGAGTCACTGCTGTGGCATTTAGTTCAAGAGTTTCAGATTTTCG 707
QY 41 LeuLeuGIyYSSerIeUAspIIeAlaProGIyAspMetLeuAspIyValIaARgARg 60
DB 708 CTTCTTGAAGTCTTTTGACATRGACACAGGTATGCTTGACAAAGTGGCAAGAGA 767
QY 61 LeuSerLeuIleYSHISProGIuCySerThraSerGIyGIyValaIIeGIuHIS 80
DB 768 CTTTCTTTAATAAATCAAGTCTCCACATGATGATGAGTGGAAAGCATAGAACAT 827
```



DB 1435 ----- 1435  
QY 241 ProlyscysProleuglyValaspilserlysgluValgylualaserileysVal 260  
DB 1436 ---TGATGTCCTTGGAGTACATATCAAAAGAGTTGAGAGCTTCATTAAGTA 1492  
QY 261 ProGlnLeuLysMetGluIle 267  
DB 1493 CCACATTAAATAATGAGATA 1513  
RESULT 10  
ADQ24627  
ID ADQ24627 standard; DNA; 2890 BP.  
XX  
AC ADQ24627;  
XX  
DT 26-AUG-2004 (first entry)  
XX  
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 7447.  
XX  
KM soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
KW ds.  
XX  
OS Homo sapiens.  
XX  
PN WO2004048938-A2.  
XX  
PD 10-JUN-2004.  
XX  
PF 26-NOV-2003; 2003WO-US038193.  
XX  
PR 26-NOV-2002; 2002US-0429739P.  
XX  
PA (PROT-) PROTEIN DESIGN LABS INC.  
XX  
PI Aziz N, Ginsburg WM, Zlotnick A;  
XX  
DR WPI; 2004-441208/41.  
XX  
PT Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.  
XX  
PS Example 2; SEQ ID NO 7447; 210pp; English.  
XX  
CC The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual,  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cytostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.  
XX  
SQ Sequence 2890 BP; 869 A; 609 C; 611 G; 789 T; 0 U; 12 Other;  
XX  
Alignment Scores:  
Pred. No.: 4,83e-126 Length: 2890  
Score: 1204.00 Matches: 239  
Percent Similarity: 89.51% Conservative: 0  
Best Local Similarity: 89.51% Mismatches: 4  
Query Match: 86.93% Indels: 24  
DB: 12 Gaps: 1  
US-10-649-273-2\_COPY\_148\_414 (1-267) x ADQ24627 (1-2890)  
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20

DB 1442 ATGAGGCTCATGACTTACTATTAGGTGACCAATTAAGTAGAATTTCTTTTACTT 1501  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuValLeuValGlnGlyValSerAspPheLeu 40  
DB 1502 CTTTGATTTCTGGAGGCTCACTGCTGTGGCACTTAGTTCAGAGATTTCGATTTTCG 1561  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
DB 1562 CTTCCTGAAAGTCTTTGGACATAGCACACAGGTACATGCTTGCACAAAGGTGGCAAGAGA 1621  
QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyLysValAlaIleGlnHis 80  
DB 1622 CTTCCTTAAATTAACATCCAGAGTGTCCACATGAGTGGGAAAGCCATAGAAACAT 1681  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
DB 1682 TTGGCCAAACAGAAATAGATTTCATTTTGACATCAAACTCCCTTCATCATCTTAA 1741  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
DB 1742 AATTGATTTTCTTTTACTGACCTTCAACACGTTACTGATTAATATATAGAAAAAG 1801  
QY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140  
DB 1802 GAAAAAGAGAGGATATGAGAAAGGGCAATCTGCTTCAGACGACGACATTGCTGCC 1861  
QY 141 ThrValGlnHisSerThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
DB 1862 ACAGTACACACACAAAGGACATGTCATCTGTGTGAAACACACATCGGGCTATTCGTTT 1921  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180  
DB 1922 TGTTAGCGAGAGAGCTTGTATCTCAAAATATATGACAGTCTGTTGATCTGTGTGTC 1981  
QY 181 AlaSerAsnPheTyrlleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200  
DB 1982 GCAAGTAACTTCTGTATCCGACAGCTCTGGAATTTTAAACAAAGCAACAGTGCAC 2041  
QY 201 LeuLeuCyProProProArgLeuCysThrAspAsnGlyIleMetIleAlaIlePheAsnGly 220  
DB 2042 TTGTTGTGTCCTCTCCACAGCTATGCACTATATAGCATTAATGATTGCA----- 2092  
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrlle 240  
DB 2092 ----- 2092  
QY 241 ProlyscysProleuglyValaspilserlysgluValgylualaserileysVal 260  
DB 2093 ---TGATGTCCTTGGAGTACATATCAAAAGAGTTGAGAGCTTCATTAAGTA 2149  
QY 261 ProGlnLeuLysMetGluIle 267  
DB 2150 CCACATTAAATAATGAGATA 2170  
RESULT 11  
ADE31345/C  
ID ADE31345 standard; DNA; 3358 BP.  
XX  
AC ADE31345;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Human diagnostic and therapeutic polynucleotide (dithp), SEQ ID No 100.  
XX  
DB diagnostic and therapeutic polynucleotide; dithp; antiarteriosclerotic;  
XX antiinflammatory; cerebroprotective; antiiphaemic; antidiabatic;  
XX immunosuppressive; neuroprotective; nootropic; neuroleptic; tranquilizer;  
XX osteopathic; antiarthritic; antineumatic; cyclostatic; hepatotropic;  
XX vitucide; haemostatic; anti-HIV; antidiyroid; lthyromimetic;  
XX dermatological; antibacterial; fungicide; antiparasitic; anticonvulsant;  
XX thrombolytic; anticoagulant; anorectic; vasotropic; antidiacer;  
XX gene therapy; protein replacement therapy; human; gene; ds.

XX Homo sapiens.  
OS  
XX  
PN MO2003062376-A2.  
XX  
XX  
PD 31-UTL-2003.  
XX  
PF 13-JAN-2003; 2003MO-US001096.  
XX  
PR 16-JAN-2002; 2002US-0349384P.  
PR 17-JAN-2002; 2002US-0349413P.  
PR 17-JAN-2002; 2002US-0349946P.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
PI Jones AL, Dahl CR, Gierzen D, Chinn J, Dufour GE, Jackson JL;  
PI Yu JY, Tuason O, Yap PE, Amshey SR, Dam TC, Liu TP, Gerslein EH;  
PI Peralta CH, Lewis SA, Chen A, Marwaha R, Lan KY, Urashka ME;  
PI Kristnam SR, Kolluru V, Panesar IS;  
XX  
DR MPI: 2003-636732/60.  
DR P-PSDB; ADE31156.  
XX  
PT New human diagnostic and therapeutic polynucleotides and polypeptides,  
PT useful for diagnosing, treating or preventing e.g. leukemia, brain  
PT cancer, atherosclerosis, AIDS, thyroiditis, infections, obesity, stroke  
PT or Alzheimer's.  
XX  
PS Claim 1; SEQ ID NO 100; 634bp; English.  
XX  
XX The invention relates to a novel isolated human diagnostic and  
CC therapeutic polynucleotide (designated dithp). The novel dithp  
CC polynucleotide comprises: any of 188 DNA sequences consisting of 195-7798  
CC base pairs fully defined in the specification; a polynucleotide  
CC comprising a naturally occurring polynucleotide sequence at least 90%  
CC identical to the dithp polynucleotide; a polynucleotide complementary to  
CC the dithp polynucleotide or its polynucleotide which is at least 90%  
CC identical; or an RNA equivalent of any of the polynucleotides mentioned  
CC above. The dithp polynucleotides have the following activities:  
CC antiatherosclerotic, antiinflammatory, cerebroprotective, antilipemic,  
CC antidiabetic, immunosuppressive, neuroprotective, nootropic, neuroleptic,  
CC tranquilizer, osteopathic, antiarthritic, antirheumatic, cystostatic,  
CC hepatoprotective, virocidic, haemostatic, anti-HIV, antihypertoid, thrombolytic,  
CC dermatological, antibacterial, fungicide, antiparasitic, anticonvulsant,  
CC thrombolytic, anticoagulant, anorectic, vasotropic, and anticulcer. The  
CC novel dithp polynucleotides polypeptide can be used in gene therapy and  
CC protein replacement therapy. The dithp polynucleotides or dithp  
CC polypeptides are useful for diagnosing, preventing or treating diseases  
CC associated with the expression of human molecules. In particular, these  
CC diseases include cancers (e.g. adenocarcinoma, leukaemia, melanoma, brain  
CC cancer, breast cancer, cervix cancer, bone cancer, liver cancer, lung  
CC cancer) or other cell proliferative disorders (e.g. arteriosclerosis,  
CC atherosclerosis, cirrhosis, hepatitis, polycythemia vera, primary  
CC thrombocytopenia), autoimmune/inflammatory disorders (e.g. AIDS,  
CC Addison's disease, thyroiditis, Crohn's disease, Graves' disease,  
CC Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid  
CC arthritis, scleroderma, systemic lupus erythematosus), infections (e.g.  
CC viral, bacterial, fungal or parasitic infection), developmental disorders  
CC (e.g. Cushing's syndrome or epilepsy), endocrine disorders (e.g.  
CC thrombosis, hypopituitarism, hypogonadism, gigantism, goiter) metabolic  
CC disorders (e.g. hypercholesterolemia, hypoglycaemia, diabetes,  
CC hyperlipidaemia, obesity), neurological disorders (e.g. ischaemic  
CC cerebrovascular disease, stroke, Alzheimer's disease, Pick's disease,  
CC Huntington's disease, Parkinson's disease, Creutzfeldt-Jakob disease,  
CC anxiety, schizophrenia), gastrointestinal disorders (e.g. ulcers),  
CC transport disorders (e.g. akinesia or multidrug resistance), or  
CC connective tissue disorders (e.g. Paget's disease or rickets). This  
CC polynucleotide sequence represents one of the human dithp DNA sequences  
CC of the invention.  
XX  
XX Sequence 3358 BP; 1105 A; 577 C; 601 G; 1075 T; 0 U; 0 Other;  
XX  
XX Alignment Scores:

Pred. No.: 2,996-102 Length: 3358  
Score: 595.50 Matches: 250  
Percent Similarity: 44.82% Conservative: 1  
Best Local Similarity: 44.64% Mismatches: 7  
Query Match: 71.88% Indels: 307  
DB: 10 Gaps: 2  
US-10-649-273-2\_COPY\_148\_414 (1-267) x ADE31345 (1-3358)  
QY 10 LeuThrAsnLysValGluPheProPheLeuValLeuLeuIleSerGlyVhiCysLeu 29  
DB 3067 GTGACCAATTAAGTGAATTTCTTTTAGTCT-TTGATTTCTGGAGTCACTGCTCG 3009  
QY 30 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 49  
DB 3008 TGGGCAATAGTTCAAGAGAGTTTCAGATTTTCTGCTTCTGGAAAGTC-TTGGACATAGCA 2950  
QY 50 ProGlyAspMet----- 53  
DB 2949 CCAGGTGACATGCGTTGCAAGGTAATTAAGATTATTTCTCATCTTTTGTATGT 2890  
QY 53 ----- 53  
DB 2889 TGTCCATTCACTAAGTAGCAATAGATGTGCTACCACTTCACTTAATTTCTGAA 2830  
QY 53 ----- 53  
DB 2829 TTTTATCTTAGTAACTGACACAAATTCATATGTGAGAAAAAATAGAAACAGTA 2770  
QY 53 ----- 53  
DB 2769 GTACAGAAATTTATATTTCTTAGCCTTTCTTAATAAATGTAGAGGTTATATCTGTA 2710  
QY 53 ----- 53  
DB 2709 CATAAAGCGTGAATAGTTTGCAATAGATGATTAATTTTCCAAATATAGTAGTGA 2650  
QY 53 ----- 53  
DB 2649 AGAAGTCCTTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2590  
QY 53 ----- 53  
DB 2589 TTAACATAGAGCAATTAAGATGCAATGACAGAAATTAATGACAAATTAATTAACCA 2530  
QY 53 ----- 53  
DB 2529 CAGACAGGATCCCCCGACCCCTTTGTTTGAATTAATTAATTAATTAATTAATTAAT 2470  
QY 53 ----- 53  
DB 2469 ATAGAAATTAAC 2410  
QY 53 ----- 53  
DB 2409 ACAAGTTCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2350  
QY 53 ----- 53  
DB 2349 CCAACCAATAGAGAGCAAAATAGACAGGGGAGTATGCTCTTATTTGTTGGGCTC 2290  
QY 53 ----- 53  
DB 2289 ATCATAGAGAAACAGGTTGCTGCTTAACCTGAATATGACTATTAATTTGCCAAA 2230  
QY 53 ----- 53  
DB 2229 GTATAGCATGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 2170  
QY 54 -----Leu 54  
DB 2169 CTTTGCAATCTTTTGTGTTTCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTT 2110

```
QY 55 AspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSerGly 74
DB 2109 GATAG-GTGGCAAGAGACTTCTTTAATTAACATCCAGAGCTCCACCATGAGTGGT 2051
QY 75 -GlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisIleAspIleLysPr 94
DB 2050 GGGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATTTCAATTTGACATCAACACC 1991
QY 94 oProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAs 114
DB 1990 TCCCTTGATCATGCTCAAAATATGATTTCTTTTACTGAGCTTCAACACGTTACTGA 1931
QY 114 pLysIleIleMetLysLysGlnLysGlnGlu-----GlyIleGlnLysG 124
DB 1930 TAAATATATATGAAAAAGAAAAAGGAAAGTATATTCTAATTAGTAAAGTTGAACA 1871
QY 125 -----GlyIleGlnLysG 129
DB 1870 GATTAATATTCCTGATGCTGCTTAAATAATGCTGCTCATTTCTGACGATTTGAGAGG 1811
QY 129 IyGlnIleLeuSerSerAlaIleAspIleAlaIleThrValGlnHisThrMetAlaCysH 149
DB 1810 GGCAATATCCTGCTTACGACGACATGCTGCTCCACAGTACACACACATGACATGCTC 1751
QY 149 ILeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProG 169
DB 1750 ATCTTGTGAAAAGAACACATCGGCTATTTCTGTTTGTAAAGAGAGAGACTTGTACTTC 1691
QY 169 IAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrlleArgArg 189
DB 1690 AAAATATATCAGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1632
QY 189 ILeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgArgLeu 209
DB 1631 CTCTGAAAATTTTAAAC-AACGCAACACAGTGCATTTGTGTGTCTCCCTCCAGACTAT 1573
QY 209 yThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuG 229
DB 1572 GCACGTATATGACATTAATGATTCATGATGATGATGATGATGATGATGATGATGATGATG 1513
QY 229 IyIleLeuHisAspIleGluGlyIleArgTyrlleArgTyrlleArgTyrlleArgTyrlle 249
DB 1512 GCATTTTACATGACATAGAGGATCGATCGATGACCAAAATGTCCTCTTGAGATAGACA 1453
QY 249 ILeuSerGlnValAlaGlyValAlaSerIleLysValProGlnLeuLysMetGlnIle 267
DB 1452 TATCAAAAGAGTTGAGAGACTTCCATTAAGTACCAAAATTAATAATGAGATTA 1397
RESULT 12
ID ABO75508 standard; DNA; 1572 BP.
XX ABO75508;
AC ABO75508;
XX ABO75508;
DT 07-NOV-2002 (first entry)
XX
DE Murine sialoglycoprotease-like gene sequence SEQ ID NO:7.
XX
XX Murine; mouse; protease; calcium activated neutral protease type 5;
KM CAPN5; tryptase 4; sialoglycoprotease; enzyme; genetic disease;
KM neurological; neuropsychological; psychotic illness; transgenic animal;
KM gene; db.
XX
OS Mus musculus.
XX
XX WO200245491-A2.
XX
XX 13-JUN-2002.
XX
XX 05-DEC-2001; 2001WO-US046405.
XX
XX 06-DEC-2000; 2000US-0251803P.
XX
XX 06-DEC-2000; 2000US-0251820P.
```

```
PR 13-DEC-2000; 2000US-0255971P.
XX
XX (DELT-) DELTAGEN INC.
PA
XX
PI Allen KD, Leviten MW;
XX
XX WPI; 2002-657389/70.
XX
XX Novel transgenic animal, comprising a disruption in protease target gene,
PT is useful for identifying agents that ameliorates a phenotype associated
PT with a disruption in a protease target gene.
XX
XX Example 3; Fig 7; 62pp; English.
XX
XX The present invention describes a non-human transgenic animal (I)
CC comprising a disruption in a protease target gene (PG) selected from
CC calcium activated neutral protease type 5 (CAPN5) gene, tryptase 4 gene
CC and sialoglycoprotease-like gene. Also described is a targeting construct
CC (II), comprising a first polynucleotide sequence homologous to at least a
CC first portion of PG, a second polynucleotide sequence homologous to at
CC least a second portion of PG and a selectable marker. (II) is useful for
CC producing a transgenic mouse comprising a disruption in a protease target
CC gene, by introducing (II) into a cell, introducing the cell into a
CC blastocyst, implanting the resulting blastocyst into a pseudopregnant
CC mouse, where the pseudopregnant mouse gives birth to a chimeric mouse,
CC and breeding the chimeric mouse to produce the transgenic mouse. (I) is
CC useful for identifying an agent that modulates the expression or function
CC of a protease target gene, by administering an agent to (I) and
CC determining whether the expression or function of the disrupted protease
CC target gene in (I) is modulated. (II) is also useful for testing the
CC efficacy of proposed genetic and pharmacological therapies for human
CC genetic diseases, such as neurological, neuropsychological or psychotic
CC illness. The present sequence represents murine sialoglycoprotease-like
CC gene sequence, which is used in an example from the present invention
XX
XX
SQ Sequence 1572 BP; 459 A; 337 C; 340 G; 429 T; 0 U; 7 Other;
XX
XX
Alignment Scores:
Pred. No.: 2 67e-101 Length: 1572
Score: 983.00 Matches: 210
Percent Similarity: 84.21% Conserved: 14
Best Local Similarity: 78.95% Mismatches: 38
Query Match: 70.97% Indels: 6
DB: Gaps: 2
US-10-649-273-2_COPY_148_414 (1-267) x ABO75508 (1-1572)
QY 1 MetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 445 ATGGAGGCTCAGCAGCATGATTAAGGCTCACCAATAAGATGATN-CTTTTAAATT 503
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 504 CTTTGAATTTCTGGCGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 563
QY 41 LeuLeuGlnLysSerLeuAspIleAlaProGlnLysMetLeuAspLysValAlaArg 60
DB 564 CTCCTTGGGAAGCTTTTGAATAGACACAGGCAACATGCTTGACAAAGTGGCAAGAA 623
QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyValAlaIleGlnHis 80
DB 624 CTTTCTTAATCAACATCCAGATGTTCTTCAATGATGATGATGATGATGATGATGATGATG 683
QY 81 LeuAlaLysGlnGlyAsnArgPheHisIleAspIleLysProProLeuHisHisAlaLys 100
DB 684 TTGGCCAAAGAGCAAAATATGATTTCAATTTTCAATCAATCCACTATGACAGATGCTAAG 743
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
DB 744 AATTCGATTTTCTTTCACGGGACTTCAACATTTACTGTATAGCTTAATTAACACACAG 803
QY 121 GluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaIleAspIleAlaAla 140
```

DB 804 GAAAAAGAGAGGCGATTGAGAGGGGCAATCTGTCTATCATGCTGACAGCATTTGCTGCT 863  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValIleuArgThrHisArgAlaIleLeuPhe 160  
DB 864 GGGGTCACACATCCACACAGGCTGCCACCTTCCGAAAAGAACACATCGCGCTATTCTGTTT 923  
QY 161 CysIleuGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 180  
DB 924 TCAGACGCAAAAATTTGGCTCTCTCCAGCTACACGCAATATTAGTTGATCTGGAGGCTT 983  
QY 181 AlaSerAspPheValIleuArgAlaIleuGlnIleuThrAsnAlaThrGlnCysThr 200  
DB 984 GGAAAGTAACTTGTACATCCCAAGCATTCGAAATTGTCGCAATGCAACGCATGCACCG 1043  
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTPAsnGly 220  
DB 1044 TTGTTGTGTCCTCT-TCAGACGTGTGCTGACATGACATGGCATGATGATTCATGGAATGGA 1102  
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeu-HisAsp-IleGluGlyIleArgTyrg 240  
DB 1103 ATTGGAAGATTACGTGCTGCTGCTGCTGCTTTACCATGATGATGAGACATTCGGTATT 1162  
QY 240 IupProIysCysProLeuGlyVal--AspIleSerIysGluValGlyGluAlaSerIle 259  
DB 1163 AACCCAAATGTCCTCTTGTGAGTGAGGCAATTATCCGAAAGATTGGCAGA---AGCTTGC 1219  
QY 259 ysaValProGlnLeu 263  
DB 1220 CCATTAAAAAGTTA 1233  
RESULT 13  
ID AAS84622 standard; cDNA; 2734 BP.  
XX AAS84622;  
AC AAS84622;  
DT 13-FEB-2002 (first entry)  
XX DNA encoding novel human diagnostic protein #20426.  
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.  
OS Homo sapiens.  
XX WO200175067-A2.  
PN 11-OCT-2001.  
PD 30-MAR-2001; 2001MO-US008631.  
XX 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX (HYSF-) HYSFQ INC.  
PA Drmanac RT, Liu C, Tang YT;  
XX P-PSDB; ABG20435.  
DR WPI: 2001-639362/73.  
XX P-PSDB; ABG20435.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX Claim 1; SEQ ID NO 20426; 103bp; English.  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed

CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 2734 BP; 773 A; 545 C; 639 G; 763 T; 0 U; 14 Other;  
Alignment Scores:  
Pred. No.: 2,61e-57 Length: 2734  
Score: 599.00 Matches: 131  
Percent Similarity: 71.57% Conservative: 10  
Best Local Similarity: 66.50% Mismatches: 22  
Query Match: 43.25% Indels: 34  
DB: 5 Gaps: 6  
US-10-649-273-2\_COPY\_148\_414 (1-267) x AAS84622 (1-2734)  
QY 71 ThrMetSerGlyIleValAlaIleGlnHisLeuAlaIleGlnGlyAsnArgPheHisPhe 90  
DB 206 ACCCTGCTAGGGGGCGCGCT-----AGATTTCAT--- 235  
QY 91 AspIleIysProProLeuHisAlaIlys-----AsnCys----- 102  
DB 236 -----CCTATTCTCCAGTGAAGATGACGCTGTGAGAACCTGCGACGA 283  
QY 103 -----AspPheSerPheThrGlyLeuGlnHisValThrAsp 114  
DB 284 TCTGCTTGTGGAGAGCCGACGAGACTTTTACTATGCGGAGGCAATTCGGAGTACG 343  
QY 115 LysIleIleMetIysGlyIleGlyIleGlu-----GlyIleGlyIleGlyIleLeu 132  
DB 344 CGTCTTACATGACGACGACGACGACGACGAGAGGGGGATGAGAGAGGGGCAATCTG 403  
QY 133 SerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValIys 152  
DB 404 TCTTCAGACGACGACATGCTGCGACACATACGACACACATGCAATGCTATCTTGAAA 463  
QY 153 ArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuProGlnAsnAla 172  
DB 464 AGAACACATCGGGCTATTCTGTTTGTGAGAGAGAGACTTTTACTTCATTAATATGCA 523  
QY 173 ValIleuValAlaSerGlyIleValAlaSerAspPheValIleuArgAlaLeuGluIle 192  
DB 524 GTACTGTTGATCTGCGTGTGCTGCCAAGTAACTTCTATATCCGACGAGCTTGGAATT 583  
QY 193 LeuThrAsnAlaThrGlnCysThrIleuLeuCysProProProArgLeuCysThrAspAsn 212  
DB 584 TTACAAACGCAACACAGTGCACCTTGTGTCTCTCCGACGACATGACATGATAT 643  
QY 213 GlyIleMetIleAlaTPAsnGlyIleGluArgLeuArgAlaGlyIleuGlyIleuHis 232  
DB 644 GGCATTATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 703  
QY 233 AspIleGluGlyIleArgTyrgIupProIys-----CysProLeuGlyVal 247  
DB 704 GACATGAGAGGCGATCGCTATGACCAAGATATGTCGCTCTTCAGAGGCTG 754  
RESULT 14  
ADL86725  
ID ADL86725 standard; DNA; 371 BP.  
XX





DB: 12 Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x ADL86726 (1-371)

```

QY 137 AspIleAlaIaIaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArg 156
    |||||
Db 2 GACATTGCTGCTGCGTACAGCAGCAACGCGTCCACCTTGGAAAAGAACATCGC 61
    |||||
QY 157 AlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAla 176
    |||||
Db 62 GCTATTCTGTTTTCAGACGACAGAAATTGCTCTCTCCAGCTAACGACGATTTAGTTGTA 121
    |||||
QY 177 SerGlyGlyValAlaAspAsnPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAla 196
    |||||
Db 122 TCTGGAGGTGTTCCAGTACTTGTACATCGAAAACATTGGAAATTGTGCAAAATGCA 181
    |||||
QY 197 ThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIle 216
    |||||
Db 182 ACCGAGTGCACGTGTGTGTCTCCAACTGCACTGCACTGCACTGCACTGCACTGCACTGAT 241
    |||||
QY 217 AlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGly 236
    |||||
Db 242 GCATTGAAATGGAATGAAATTAAGTACGTCGAGCTTNGGCNTTNTTACNTGANNNTAGAAAGAC 301
    |||||
QY 237 IleArgTyrGluProLysCysProLeuGlyValAspIleSerLysGluValGlyValAla 256
    |||||
Db 302 ATCCNNTATGNAACCAAAATNTCNTNTTGGAGTAGACATNTCCAGAGAAATTGCAGAAAGCT 361
    |||||
QY 257 SerIleLys 259
    |||||
Db 362 GCCATATAAA 370
    |||||

```

Search completed: June 16, 2005, 20:15:08  
 Job time : 396.989 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 16, 2005, 18:30:49 ; Search time 132.629 Seconds  
(without alignments)  
3294.036 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_148\_414

Perfect score: 1385  
Sequence: 1 MEAHALTRLNKVFPPFLV.....DISKEVGASIKVPOLKMEI 267

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/US10649273/runat\_15062005\_111417\_6057/app\_query.fasta\_1.1429  
-DB=Issued Patents NA -QPMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdd  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTPM=ptco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10649273\_QCEN\_1.1.177\_@runat\_15062005\_111417\_6057 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.\*

1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length  | DB ID              | Description       |
|------------|--------|-------------|---------|--------------------|-------------------|
| 1          | 1385   | 100.0       | 1416    | US-09-774-528-177  | Sequence 177, App |
| 2          | 1385   | 100.0       | 1526    | US-10-067-443-23   | Sequence 23, Appl |
| 3          | 1385   | 100.0       | 2197    | US-10-067-443-1    | Sequence 21, Appl |
| 4          | 1358.5 | 98.1        | 1387    | US-10-067-443-21   | Sequence 20, Appl |
| 5          | 824.5  | 59.5        | 14364   | US-10-067-443-20   | Sequence 38, Appl |
| 6          | 308.5  | 22.3        | 94750   | US-09-536-002-38   | Sequence 806, App |
| 7          | 307    | 22.2        | 1053    | US-09-540-236-806  | Sequence 884, App |
| 8          | 291.5  | 21.0        | 1059    | US-09-252-991A-884 | Sequence 801, App |
| 9          | 291.5  | 21.0        | 1206    | US-09-252-991A-801 | Sequence 1, Appl  |
| 10         | 289.5  | 20.9        | 1830121 | US-09-557-884-1    | Sequence 1, Appl  |
| 11         | 289.5  | 20.9        | 1830121 | US-09-643-990A-1   | Sequence 1, Appl  |
| 12         | 282    | 20.4        | 996     | US-09-902-540-6612 | Sequence 6612, Ap |

|      |       |      |         |   |                     |                   |
|------|-------|------|---------|---|---------------------|-------------------|
| C 13 | 282   | 20.4 | 2582    | 4 | US-09-902-540-503   | Sequence 503, App |
| 14   | 280.5 | 20.3 | 1375    | 4 | US-09-543-681A-2341 | Sequence 2341, Ap |
| 15   | 270.5 | 19.5 | 1014    | 1 | US-08-087-797-1     | Sequence 1, Appl  |
| 16   | 268.5 | 19.4 | 1032    | 4 | US-09-489-039A-2050 | Sequence 2050, Ap |
| 17   | 261   | 18.8 | 1092    | 4 | US-09-107-532A-2955 | Sequence 2955, Ap |
| 18   | 254   | 18.3 | 1008    | 3 | US-08-987-121A-5    | Sequence 5, Appl  |
| 19   | 254   | 18.3 | 1011    | 3 | US-08-987-121A-3    | Sequence 3, Appl  |
| 20   | 250   | 18.1 | 1006    | 3 | US-08-961-083-51    | Sequence 51, Appl |
| 21   | 250   | 18.1 | 1006    | 4 | US-09-536-784-51    | Sequence 51, Appl |
| 22   | 250   | 18.1 | 10974   | 3 | US-08-961-527-214   | Sequence 214, App |
| 23   | 248   | 17.9 | 1011    | 3 | US-09-066-512-1     | Sequence 1, Appl  |
| 24   | 246   | 17.8 | 1011    | 4 | US-09-583-110-2196  | Sequence 2196, Ap |
| 25   | 241   | 17.4 | 1011    | 4 | US-09-107-433-1618  | Sequence 1618, Ap |
| 26   | 236.5 | 17.1 | 640681  | 4 | US-09-790-988-1     | Sequence 1, Appl  |
| 27   | 236   | 17.0 | 822     | 4 | US-09-710-279-727   | Sequence 727, App |
| 28   | 236   | 17.0 | 1101    | 4 | US-09-134-000C-1551 | Sequence 1551, Ap |
| 29   | 236   | 17.0 | 3993    | 4 | US-09-710-279-3985  | Sequence 3985, Ap |
| 30   | 233   | 16.8 | 1107    | 3 | US-09-134-001C-1072 | Sequence 1072, Ap |
| 31   | 220   | 15.9 | 1026    | 3 | US-09-149-624-1     | Sequence 1, Appl  |
| 32   | 218   | 15.7 | 3064    | 3 | US-09-221-017B-794  | Sequence 794, App |
| 33   | 218   | 15.7 | 123025  | 4 | US-09-198-452A-1    | Sequence 1, Appl  |
| 34   | 218   | 15.7 | 1230230 | 4 | US-09-438-185A-1    | Sequence 102, App |
| 35   | 216   | 15.6 | 15249   | 4 | US-08-956-171E-102  | Sequence 102, App |
| 36   | 216   | 15.6 | 15249   | 4 | US-08-781-966A-102  | Sequence 102, App |
| 37   | 215.5 | 15.6 | 4403765 | 3 | US-09-103-840A-2    | Sequence 2, Appl  |
| 38   | 215.5 | 15.6 | 4411529 | 3 | US-09-103-840A-1    | Sequence 1, Appl  |
| 39   | 211.5 | 15.3 | 1155    | 4 | US-09-602-777A-147  | Sequence 147, App |
| 40   | 207.5 | 15.0 | 1664976 | 4 | US-08-916-421B-1    | Sequence 1, Appl  |
| 41   | 207.5 | 15.0 | 1664976 | 4 | US-09-692-570-1     | Sequence 1, Appl  |
| 42   | 192.5 | 13.9 | 1663    | 4 | US-09-620-312D-6    | Sequence 6, Appl  |
| 43   | 185.5 | 13.4 | 42325   | 4 | US-08-311-731A-131  | Sequence 131, App |
| 44   | 181.5 | 13.1 | 432     | 4 | US-09-328-352-261   | Sequence 261, App |
| 45   | 169   | 12.2 | 876     | 4 | US-09-724-623-19    | Sequence 19, Appl |

#### ALIGNMENTS

##### RESULT 1

US-09-774-528-177  
; Sequence 177, Application US/09774528

; Patent No. 6743619  
; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom  
; APPLICANT: Goodrich, Ryle

; APPLICANT: Lin, Chenghua  
; APPLICANT: Asundi, Vinod

; APPLICANT: Ren, Feiyan  
; APPLICANT: Zhang, Jie

; APPLICANT: Zhao, Qing A.  
; APPLICANT: Yang, Yonghong

; APPLICANT: Xue, Aidong J.  
; APPLICANT: Wehrman, Tom

; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Wang, Dunrui

; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: Polypeptides

; FILE REFERENCE: 802  
; CURRENT APPLICATION NUMBER: US/09-774,528

; NUMBER OF SEQ ID NOS: 441  
; SOFTWARE: pt\_fl\_genes Version 2.0

; SEQ ID NO 177  
; LENGTH: 1416

; TYPE: DNA  
; ORGANISM: Homo sapiens

; FEATURE:  
; NAME/KEY: CDS

; LOCATION: (205)..(1305)  
; US-09-774-528-177

Alignment Scores:

Pred. No.: 2,516-177 Length: 1416  
 Score: 1385.00 Matches: 267  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 4 Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-09-774-528-177 (1-1416)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
 DB 502 ATGGAGGCTCATGCACTTACTATTAGGTTGACCAATTAAGTAGAATTTCTTTTAAAGT 561  
 QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
 DB 562 CTTTTCATTTCTGGAGGCTCATCTGCTGTTGGCATTTAGTTCAAGGAGTTTCAGATTTCTG 621  
 QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
 DB 622 CTTCTTGGAAAGCTTTGGCATAGCAACGAGTGACATGCTTGAACAAGGTGGCAAGAAGA 681  
 QY 61 LeuSerLeuIleLysHisProGlyLysSerThrMetSerGlyGlyLysAlaIleGlnHis 80  
 DB 682 CTTTCTTATATATAAATCAATCCAGAGCTCTCCACATGATGAGTGAGGAAAGCCATAGAACAT 741  
 QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProPheLeuHisAlaLys 100  
 DB 742 TTGGCCAAACAGGAATATAGATTTTCATTTTGACATCAAACTCCCTTGCATCATGCTTAA 801  
 QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
 DB 802 AATTGATTTTCTTTTACTGACCTTCAACACGTTACTGATTAATTAATTAAGAAAAAG 861  
 QY 121 GluLysGlnGlyLysIleGlnLysGlnIleLeuSerSerAlaAlaAspIleAlaAla 140  
 DB 862 GAAAAAGAGAAAGTTGAGAAAGGGGCAATCTGCTTCCACAGCAGCATTTGCTGCC 921  
 QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
 DB 922 ACGATACAGACACAAATGGCATGTCATCTTTGTAAGAAGACACATCGGCTATTTCTGTT 981  
 QY 161 CysLysGlnArgAspLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180  
 DB 982 TGTAAACAGAGACATTTGTAACCTCAAAATTAATGACGTACTGGTGCATCTGGGTGTC 1041  
 QY 181 AlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200  
 DB 1042 GCAAGTAACCTTCTATATCCGACAGGCTCTGGAATTTTAAACAAACGACACAGTGCAC 1101  
 QY 201 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaTPAsnGly 220  
 DB 1102 TTGTTGTGTCCTCTCCAGACTATGACATGATTAATGCAATTAATGATTAATGAAAGT 1161  
 QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGlu 240  
 DB 1162 ATTGAAGAAGCTACGTCCTGCTGGCATTTTACATGACATAGAGGCAATCCGTAAGAA 1221  
 QY 241 ProLysCysProLeuGlyValAspIleSerLysGlnValGlyGlnAlaSerIleLysVal 260  
 DB 1222 CCAAAATGTCCTTCTGAGTAGACATATCAAAAGAGTTGAGAGAGCTTCATTAAGTA 1281  
 QY 261 ProGlnLeuLysMetGluIle 267  
 DB 1282 CCACAAATTAATAATGAGATA 1302

RESULT 2  
US-10-067-443-23

; Sequence 23, Application US/10067443  
 ; Patent No. 6642041  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bristol-Myers Squibb Company  
 ; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
 ; TITLE OF INVENTION: SPINAL CORD, MP-1

; FILE REFERENCE: D0073 NP  
 ; CURRENT APPLICATION NUMBER: US/10/067,443  
 ; CURRENT FILING DATE: 2002-02-05  
 ; PRIOR APPLICATION NUMBER: US 60/266,518  
 ; PRIOR FILING DATE: 2001-02-05  
 ; PRIOR APPLICATION NUMBER: US 60/282,814  
 ; PRIOR FILING DATE: 2001-04-10  
 ; NUMBER OF SEQ ID NOS: 71  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 23  
 ; LENGTH: 1526  
 ; TYPE: DNA  
 ; ORGANISM: homo sapiens  
 ; US-10-067-443-23

Alignment Scores:

Pred. No.: 2,856-177 Length: 1526  
 Score: 1385.00 Matches: 267  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 4 Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-067-443-23 (1-1526)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
 DB 1 ATGGAGGCTCATGCACTTACTATTAGGTTGACCAATTAAGTAGAATTTCTTTTAAAGT 60  
 QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
 DB 61 CTTTTCATTTCTGGAGGCTCATCTGCTGTTGGCATTTAGTTCAAGGAGTTTCAGATTTCTG 120  
 QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
 DB 61 CTTTGAATTTCTGGAGGCTCATCTGCTGTTGGCATTTAGTTCAAGGAGTTTCAGATTTCTG 120  
 QY 121 GluLysGlnGlyLysIleGlnLysGlnIleLeuSerSerAlaAlaAspIleAlaAla 180  
 DB 121 CTTTGGAAAGCTTTTGGCATAGCAACGAGTGACATGCTTGAACAAGGTGGCAAGAAGA 180  
 QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
 DB 181 CTTTCTTATATATAAATCAATCCAGAGTCTCCACATGATGAGTGAGGAAAGCCATAGAACAT 240  
 QY 161 CysLysGlnArgAspLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180  
 DB 181 GAAAAAGAGAAAGTTGAGAAAGGGGCAATCTGCTTCTGAGACAGACATTTGCTGCC 420  
 QY 241 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
 DB 421 ACAATACAGACACAAATGGCATGTCATCTTGTGAAGAAGACACATCGGCTATTTCTGTT 480  
 QY 421 CysLysGlnArgAspLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180  
 DB 481 TGTAAACAGAGACATTTGTAACCTCAAAATTAATGACGTACTGGTGCATCTGGGTGTC 540  
 QY 181 AlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200  
 DB 541 GCAAGTAACCTTCTATATCCGACAGGCTCTGGAATTTTAAACAAACGACACAGTGCAC 600  
 QY 201 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaTPAsnGly 220  
 DB 601 TTGTTGTGTCCTCTCCAGACTATGACATGATTAATGCAATTAATGATTAATGAAAGT 660  
 QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGlu 240  
 DB 661 ATTGAAGAAGCTACGTCCTGCTGGCATTTTACATGACATAGAGGATCGGATAGAA 720

QY 241 ProlyscysProleuglyValAspIleSerlySgluValGlyGluAlaSerIlelyVal 260  
DB 721 CCAAAATGCTCTTGAGAGTACATATCAAAAGAGTGGAGAGCTTCCATTAAGTA 780  
QY 261 ProGlnleuylsMetGluIle 267  
DB 781 CCACAAATTAATAATGAGATA 801

RESULT 3  
US-10-067-443-1  
Sequence 1, Application US/10067443  
Patent No. 6642041  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN  
TITLE OF INVENTION: SPINAL CORD, MP-1  
FILE REFERENCE: D0073 NP  
CURRENT APPLICATION NUMBER: US/10/067,443  
CURRENT FILING DATE: 2002-02-05  
PRIOR APPLICATION NUMBER: US 60/266,518  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: US 60/282,814  
PRIOR FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1  
LENGTH: 2197  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (231)..(1472)  
US-10-067-443-1

Alignment Scores:  
Pred. No.: 5.23e-177 Length: 2197  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-067-443-1 (1-2197)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnlyValGluPheProPheLeuVal 20  
DB 672 ATGAGAGCTCATGCACTTACTATTAGTGACCAATTAAGTAGAATTTCTTTTACTT 731  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
DB 732 CTTTGTATTTCTGGAGGCTGCTCTTGGCATTAGTTCAAGAGTTTCAGATTTTCTG 791  
QY 41 LeuLeuGlylySerLeuAspIleAlaProGlyAspMetLeuAspLyValAlaArgArg 60  
DB 792 CTTCTTGGAANGCTTTTGACATAGCACAGTCACTCTTGACACAGGTGGCAAGAGA 851  
QY 61 LeuSerLeuIlelyHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 80  
DB 852 CTTTCTTTAATTAACATCCAGAGTGTCCACCATGAGTGGGGAAGGCCATAGACAT 911  
QY 81 LeuAlaLySgInGlyAsnArgPheHisPheAspIlelyProProLeuHisAlaLyS 100  
DB 912 TTGGCCCAACAAAGAAATAGATTTCATTTCATCAATCAAACTCCCTTCATCATCTAA 971  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLySleIleMetLyS 120  
DB 972 AATTGTATTTCTCTTACTGACTTCAACCGTTACTGATTAATAATATGAAAG 1031  
QY 121 GluLySgInGlyIleGlyLySgIlyGlnIleLeuSerSerAlaAlaAspIleAla 140  
DB 1032 GAAAGAGAGAGATTTGAGAGGCAAAATCTGTCTTGACGACGACATTCCTCC 1091  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuVallySarGthrHisArgAlaIleLeuPhe 160

DB 1092 ACAGTACGACACAAATGGCATGTCATCTTGTGAAAAGAACATCGGCTATTTCTGTT 1151  
QY 161 CyslySgInArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 180  
DB 1152 TGTAAAGCAGAGAGACTTGTACTTCAAAATATAGTACAGTACTGTGATCTGTGTGC 1211  
QY 181 AlaSerAspPheThrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200  
DB 1212 GCAAGTAACTTCTATATCCGAGAGCTCTGGAAATTTTAACAAAGCAACAGAGCACT 1271  
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGly 220  
DB 1272 TTGTGTCTCTCTCCCAAGCTATGACATGATTAATGCAATTATGATGACATGAGATGT 1331  
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlyGlyIleArgTyArg 240  
DB 1332 ATTGAAGACTACAGTCTGCGCTTGGCATTTTACATGACATAGAGGCAATCCGCTATGA 1391  
QY 241 ProlyscysProleuglyValAspIleSerlySgluValGlyGluAlaSerIlelyVal 260  
DB 1392 CCAAAATGCTCTTGAGAGTACATATCAAAAGAGTTGGAGAGCTTCCATAAAGTA 1451  
QY 261 ProGlnleuylsMetGluIle 267  
DB 1452 CCACAAATTAATAATGAGATA 1472

RESULT 4  
US-10-067-443-21  
Sequence 21, Application US/10067443  
Patent No. 6642041  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN  
TITLE OF INVENTION: SPINAL CORD, MP-1  
FILE REFERENCE: D0073 NP  
CURRENT APPLICATION NUMBER: US/10/067,443  
CURRENT FILING DATE: 2002-02-05  
PRIOR APPLICATION NUMBER: US 60/266,518  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: US 60/282,814  
PRIOR FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 21  
LENGTH: 1387  
TYPE: DNA  
ORGANISM: homo sapiens  
US-10-067-443-21

Alignment Scores:  
Pred. No.: 9.58e-174 Length: 1387  
Score: 1358.50 Matches: 266  
Percent Similarity: 91.10% Conservative: 0  
Best Local Similarity: 91.10% Mismatches: 1  
Query Match: 98.09% Indels: 25  
DB: Gaps: 1

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-067-443-21 (1-1387)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnlyValGluPheProPheLeuVal 20  
DB 465 ATGAGAGCTCATGCACTTACTATTAGTGACCAATTAAGTAGAATTTCTTTTACTT 524  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
DB 525 CTTTGTATTTCTGGAGGCTACGTCTCTTGGAATTAAGTTCAAGAGTTTCAGATTTTCTG 584  
QY 41 LeuLeuGlylySerLeuAspIleAlaProGlyAspMetLeuAspLyValAlaArgArg 60  
DB 585 CTTCTTGGAAGTCTTTTGACATAGCACAGTGAATCTTGACAAAGTGGCAAGAGA 644  
QY 61 LeuSerLeuIlelyHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 80

Db 645 CTTTCTTATATAAATCCAGAGTCTCCACCATGAGTGTGGAAAGCCATGAGCAT 704  
Qy 81 LeuAlaLySGInGlyAsnArgPheHisPheAspIleYsProPheUhiShiSaIAlYs 100  
Db 705 TTGGCCAAACAGAAATAGATTCTTTCATTTGACATCAAACTCCCTTGACATGCTAAA 764  
Qy 101 AsnCyAspPheSerPheThrgIyLeuGInHisValThraPlySileIleMetYsLys 120  
Db 765 AATGTGATTTTCTTTTACTGACCTCAACAGTTACTATTAATAATATGAAAAAG 824  
Qy 121 GluYsGluGluGluYle----- 126  
Db 825 GAAAGAGAGAAAGTATATTCTTAATTAGTAAAGTTGAACAGATTAATATTCCTGATG 884  
Qy 127 -----GluYsGluGluGluIleLeuSerSerYsAla 135  
Db 885 TGCCTTAAAAATAGCTGCTCATTTTCTGACGATAGAGAAAGGAGAAATCCGTCTTCAGCA 944  
Qy 136 AlaAspIleAlaAlaThraValGInHisThrMetAlaCySHisLeuValYsArgThriS 155  
Db 945 GAGACATTTGCTGCCACAGTACAGCACATGSCATGTCATCTTGGAAGAAAGACAT 1004  
Qy 156 ArgAlaIleLeuPheCySlySGInArgAspLeuLeuProGlnAsnAsnAlaValLeuVal 175  
Db 1005 CCGGCTATTCTGTTTGTATGACAGAGACTGTTGTTACTCAAAATATATGACGACTGTT 1064  
Qy 176 AlaSerGlyGlyValAlaSerAsnPheTrIleArgArgAlaLeuGluIleLeuThraSn 195  
Db 1065 GATCTGGTGGTGTCCGAAAGTAACTTCTATATCCGAGACTCTCGAAATTTTAAACAAC 1124  
Qy 196 AlaThrgInCySthrLeuLeuCySProProArgLeuCySThraPAsnGlyIleMet 215  
Db 1125 GCAACACAGTGCCTTTGTTGTCTCTCCCTCCAGACTATGCACTGTAATGGCATTAATG 1184  
Qy 216 IleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGluIleLeuHISAspIleGlu 235  
Db 1185 ATTCGATGGAATGTATTGAAGACTACGTGTGGCTTGGCATTTTACATGACATAGAA 1244  
Qy 236 GlyIleArgTrpGluProIyCySProLeuGlyValAlaSpIleSerYsGluValGlyGlu 255  
Db 1245 GGCATCCGCTATACCAAAATGCTCTTGAGTAGACATATCAAAAGAGTTGAGAA 1304  
Qy 256 AlaSerIleYsValProGlnLeuYsMetGluIle 267  
Db 1305 GCTTCATATAAGTACCAATTAATAATGAGATA 1340  
RESULT 5  
US-10-067-443-20  
; Sequence 20, Application US/10067443  
; Patent No. 6642041  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
; FILE OF INVENTION: SPINAL CORD, MP-1  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 20  
; LENGTH: 14364  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-067-443-20  
Alignment Scores:  
Pred. No.: 1,34e-99 Length: 14364  
Score: 824.50 Matches: 240

Percent Similarity: 26.73% Conservative: 0  
Best Local Similarity: 26.73% Mismatches: 2  
Query Match: 59.53% Indels: 657  
DB: 4 Gaps: 4  
US-10-649-273-2\_copy\_148\_414 (1-267) x US-10-067-443-20 (1-14364)  
Qy 1 MetGluAlaHisAlaLeuThrIleArgLeuThraSlyValGluPheProPheLeuVal 20  
Db 10845 ATGAGGCTCATGCACTTACTATAGGTTGACCAATTAAGTAGAATTTCTTTTGTAGTT 10904  
Qy 21 LeuLeuIleSerGlyGlyHisCySLeuLeuAlaLeuValGInGlyValIleSerAspPheLeu 40  
Db 10905 CTTTGAATTTCTGAGAGTCACTGCTGTGTGGCATTAGTCAAGAGTTTCAGATTTCTG 10964  
Qy 41 LeuLeuGlyYsSerLeuAspIleAlaProGlyAspMetLeuAspYsVal----- 57  
Db 10965 CTTATGGAAGGCTTTTGACATATGACACAGTGCATGCTTGACAGAGT-AAATTAAGAA 11023  
Qy 57 ----- 57  
Db 11024 TTAATTTCTCATTTCTTTTGTATGTGTCATTTCACTAAGTAGCAATPAGATGTC 11083  
Qy 57 ----- 57  
Db 11084 TACACACATTCACCTAAATATTTCTGAATTTTATCTTAGTAAAGTAAAAAATTCACAT 11143  
Qy 57 ----- 57  
Db 11144 ATGTGAGAAAAAATAGAAAGTAGTAGACAAATTTATTAATCTTAGCTTTCTTAATA 11203  
Qy 57 ----- 57  
Db 11204 AAATGTAAAGGTTTCATATCTGTATCAATAAAGCTGAATAGTTTGCAGATACACTTAATG 11263  
Qy 57 ----- 57  
Db 11264 TATTTGCCAAATATATGATGTAAAGAAAGTGTCTGTAAACTAACATACCTGCAAAAAA 11323  
Qy 57 ----- 57  
Db 11324 GGTAAATAGAGAT 11383  
Qy 57 ----- 57  
Db 11384 TTAATACACAAATTACTTACACACAGAGAGGTCCTCCCTCCCTCTTTGTTTAG 11443  
Qy 57 ----- 57  
Db 11444 AATATACAGAGGCTACTGCAATATATAGAAAAAACAACAACAACAACAACAACAACA 11503  
Qy 57 ----- 57  
Db 11504 CTTCCACAGTGAATATATAGAAATATAGCAAGTCTTATATATGACGTTATCAT 11563  
Qy 57 ----- 57  
Db 11564 AAGCAGTTATGCACTTCAAGCCATTTTCCAAACATATGAAAGCAACAACATAGACAG 11623  
Qy 57 ----- 57  
Db 11624 GGGCAGATTTGGCTCTTATATGTTGGGTCATATAGAAAGAGAGGTTGTCTGTAC 11683  
Qy 57 ----- 57  
Db 11684 TGAATATCAGTAT 11743  
Qy 57 ----- 57  
Db 11744 TTTTGTGTGTAT 11803  
Qy 58 -----AlaArgLeuSerLeuIle 64

```

Db 11804 AATTATATGACCTGAAAAATATGTTCTTGTAGAGTGGCAGAGACCTTTCTTAATA 11863
QY 65 LysH1sProGluCyseSerThmetSerGlyValylallegluH1sleuAlalysGln 84
Db 11864 AAACATCCAGAGTCTCCACCATGAGTGGGAAACCATAGAACCTTTGGCCAAACA 11923
QY 85 GlyAsnArgPheH1sPheAsp11elysProleuH1sh1salalysAsnCyseAspPhe 104
Db 11924 GGAATATGATTTCTATTTGACATCAACCTCCCTTGATCATGCTTAAAAATTTGATTTT 11983
QY 105 SerPheThrglyleuGlnH1svalThrAspLys11lelmetelysGlyValylGlu 124
Db 11984 TCTTTTACGAGACTCAACAGCTTACATGATTAATGAAAAAGAAAAAGAGAA 12043
QY 124 ----- 124
Db 12044 GGATATTTTCTAATTAAGTAAAGTGAACAGATAAATATCTCGAGTTGCTGCTAAAAATA 12103
QY 125 -----Gly11leglylysglyGln11leusSerSerAlalAsp11eal 139
Db 12104 GCTGCTCATTTCTGCAGGTATTTGAGAAAGGCAATCTGCTTCCAGCAGACGATTC 12163
QY 139 AlaThrValGlnH1sThmetAlaCysh1sleuValylsArgThrh1sAlGAlal1ele 159
Db 12164 TGGCAGAGTACAGACACAAATGGCATGTCTTGTGAAAAAGAACATCGGCTATTCT 12223
QY 159 uPheCyshysGlyAspArgPheleuProGlnAsnAsnAlalValleu----- 174
Db 12224 GTTTTGTAAACAGAGAGACTTGTTAACCTCAAAATTAAGCATGACTGTAAAGTTTATCTC 12283
QY 174 ----- 174
Db 12284 ATTTTATAGTAATAGTACACTTTCGAATATGTTACTTTTCCAGACCTTGACCTTG 12343
QY 174 ----- 174
Db 12344 TGTTTAGATGAACAGATCTTTATGCTTATGCTAGTACCCTGACAGTAAATATATGAG 12403
QY 174 ----- 174
Db 12404 GATAGAAAGACTAAACAGCCATTTCTGTACTAGTTGGTAGCTTTATGAGACAGCTGTA 12463
QY 174 ----- 174
Db 12464 TAGCTTCTATGCACTAATAGTCTTATTTGATCTTCTTGTGATTTAAAAAGGGCTT 12523
QY 174 ----- 174
Db 12524 ACATATAAGAAAGTAATGAGTAAGCTATCACTATTTTAAAGAAAAATAGTGATTT 12583
QY 174 ----- 174
Db 12584 CTTTCATCTTTGATGAATATCCCTTGTGTTGTTTAAATTAAGCCAGTCAATTT 12643
QY 174 ----- 174
Db 12644 AGCAGTGGAGNGTATTCACACTTCGAGACATAATGTTGATTAAGTCTGATTAATC 12703
QY 174 ----- 174
Db 12704 CACTATATTTGACAGCAAAATCCCTTAATTTGCTTAAAGCCTTGACAAACATCT 12763
QY 174 ----- 174
Db 12764 GTTAACTGTATCTTAACCTTATTCATTTAAAAATTAATACTAAAGTGGAAAAATGT 12823
QY 174 ----- 174
Db 12824 TAAATGTATATTCATATAGATGAATTTTACATGATATCAAAAGATATTTTTCAGA 12883
QY 174 ----- 174
Db 12884 GTTATGTATGAATATGCAAAATTAATTAATAATTCAGGGTCTTAATATAGTACTATGA 12943

```

```

QY 174 ----- 174
Db 12944 TTGAATATATTAATTAATTAATTTAGATGAAGGTGGAGAAAAATATACAAAATGCT 13003
QY 174 ----- 174
Db 13004 AGTAATGTTGTGATGCTATTAAGAAATTAATTAATTTTCTTCCAAATTTTATTAAC 13063
QY 174 ----- 174
Db 13064 ATAGATATGATCTGCCCATTTACCATCTGAAAATGGATAGTTATTTATTTATG 13123
QY 174 ----- 174
Db 13124 CTGATATTTTCTCCAGGTTAATTAAGAGCTTGCTCATATTCATATATGATTAAT 13183
QY 175 -----Val11sSerGlyVal11AsnAsnPheTyrl1eAr 187
Db 13184 TTGGTTTCTCAATTCCTTCAGGTTGCATCTGGTGTGCGAAGTAACTTATATATCG 13243
QY 187 GARGAlaleuGlu11eleuThrasnAlaThrGlnCyshThreleuCyseProProAr 207
Db 13244 CAGAGCTCGAATTTTAACAAAGCAACAGACAGTGCCTTTGTTGTCTCTCCAG 13303
QY 207 gLeuCyshrasPangly11emec11eAlarP----- 218
Db 13304 ACTATGACATGATATGSCATTATGATTCATGTAGACCAAGATATAGGTCTCAC 13363
QY 218 ----- 218
Db 13364 TCATACATATGTAATTAATTAATTCGCTTATTCATACACTTAAGCTTCTTCAGATCT 13423
QY 219 -----AsnGly11leglyuArgLeuArgA 226
Db 13424 TGGAGCTATGATTTTAATTAATTAATGCTTCTTAATTAAGAAATGATGAAGACTACGTG 13483
QY 226 laGlyleuGly11leuH1sAsp11leglyGly11eArgTyrsUpProlys 242
Db 13484 CTGGCTTGGGCTTTTACATGACATAGAAAGGATCCGCTATGAAACCAAG 13533

RESULT 6
US-09-596-002-38
; Sequence 38, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Paterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 38
; LENGTH: 94750
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc. feature
; OTHER INFORMATION: incyte template ID No. 6632636 38
; PUBLICATION INFORMATION:
; US-09-596-002-38

Alignment Scores:
Pred. No.: 3,21e-28 Length: 94750
Score: 308.50 Matches: 87
Percent Similarity: 49.61% Conservative: 40
Best Local Similarity: 33.98% Mismatches: 101
Query Match: 22.27% Indels: 29

```

DB: 4 Gaps: 8

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-09-596-002-38 (1-94750)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThr-----AsnLysValGluPheProPhe 18  
|||||  
24215 ATGGAAGGACATCTGTTAGGACCGCTATTGGCCAGATGATGACCCCAAGCTTCCGTTT 24274

QY 19 LeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAsp 38  
|||||  
24275 GGTGTCCTGCTGTGTGTCGGGGGCTATACATGCTGCTGGCCGTCGATGGTGTGGCGCTG 24334

QY 39 PheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAla 58  
:::  
24335 TATCAGATATTGGGCGAGCTTATCGATGATGATGCGGGGTGAATGCTTTGATTAACGGCA 24394

QY 59 ArgArgLeuSerLeuLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIle 78  
:::  
24395 AAAATGCTCAAACTG---CCCTATCTC-----GGTGGCCCAAAATATC 24433

QY 79 GluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProLeuHisHis 98  
|||||  
24434 GAAAAATTGACCAAAAACGGCAACCCACACCCCTGTGAGCTCCAAAGAACCCATGACGAT 24493

QY 99 AlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMet 118  
---  
24494 ---AAAGCGCTGATTTTTCGTTAGTGCATGAAACCCGCAATCATATCATCTCAAA 24550

QY 119 LysLysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIle 138  
:::  
24551 GACACACCAAAACGCCCAAGGACGCC-----GCCACAGACGACGATC 24595

QY 139 AlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIle 158  
|||||  
24596 GCGCGAAGCTTGATGATGCGGTGATGATCTTGTGTCAAAAAATGACCAAAAGCACTA 24655

QY 159 LeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaValLeuValAlaSerGly 178  
|||  
24656 CAGATGACAGGCACTTCGCCAG-----CTGGTGTGCGCAGGG 24691

QY 179 GlyValAlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGln 198  
|||||  
24692 GCGGTCTCTGCCAATCAGATGCTACGCCGACCCCTGACCGAAGCGTCCGCCAAATCGAT 24751

QY 199 CysThrLeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrp 218  
:::  
24752 GCGTCGATGACTATGATGCCCGACGAGCTATGACAGGATTAATGTCGATGATCGCTAT 24811

QY 219 AsnGlyIleGluLysArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArg 238  
|||||  
24812 GCTGCTTTTGTGGCTTCAGCTGTGACAGTGC-----GATGACTCTGGC-GGTTGCGTG 24864

QY 239 TyrGluProLysCysProLeuGlyValAspIleSerLysGluValGly 254  
|||  
24865 TATTC-----CCGATGGGATATGACACGCTTGCGCTATCGCC 24903

RESULT 7  
US-09-540-236-806  
Sequence 806, Application US/09540236  
Patent No. 6673910  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAHALIS  
FILE REFERENCE: 2709-2005-001  
CURRENT APPLICATION NUMBER: US/09/540,236  
CURRENT FILING DATE: 2000-04-04  
NUMBER OF SEQ ID NOS: 3840  
SEQ ID NO 806  
LENGTH: 1053  
TYPE: DNA  
ORGANISM: M.catahhalis  
US-09-540-236-806

[illegible]

; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 884  
 ; LENGTH: 1059  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-884

## Alignment Scores:

| Pred. No.:             | 3,67e-29 | Length:       | 1059 |
|------------------------|----------|---------------|------|
| Score:                 | 291.50   | Matches:      | 87   |
| Percent Similarity:    | 50.41%   | Conservative: | 37   |
| Best Local Similarity: | 35.37%   | Mismatches:   | 99   |
| Query Match:           | 21.05%   | Indels:       | 23   |
|                        |          | Gaps:         | 9    |

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-09-252-991A-884 (1-1059)

```

OY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys---ValGluPheProPheLeu 19
DB 711 ATGGAAGGACACCTGCGCGCCGATGCTGGAAGACGACCGCGGTTCCCGTTCGTC 652
OY 20 ValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPhe 39
DB 651 GCCTTGCTGGTTTCCCGGCTGACACCCAGTGTGCGGGTGAAGGATATCGCGCTAC 592
OY 40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 59
DB 591 CAGTTGCTTGGCAATCGGTGACGATCGCGCGGCAACCTTCGACAAAGCCGCAAG 532
OY 60 ArgLeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyValAlaIleGlu 79
DB 531 CTGATCGGCTG---GGCTATGCC-----GGTGTCCGGAATATGCC 493
OY 80 HisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProPheLysHisAla 99
DB 492 CCGCTGGCGGAGCGCGGCACTCTGCGCGGCTGTGTTCCCGCGGCGAATGCGCATCGC 433
OY 100 LysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLys 119
DB 432 CCGCGGCTGACTTCAAGCTTCAAGCGGCTCAAGACCTTACCTGAAC---ACCTGCGAG 376
OY 120 LysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 139
DB 375 CGTTGCGTGAAGCCCGCGGACGACGACGACGACGACGACGACGACGACGACGACGACG 325
OY 140 AlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeu 159
DB 324 CTGGCGCTTCACAGCCCGGCGGTGTCAGACCTGCTGATCAAGTCCGCGCGCTTG--- 268
OY 160 PheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGly 179
DB 267 -----AAGCAGACCGGCTG---AAGAAC-----CTGTGATCGCGCGGCGT 229
OY 180 ValAlaSerAspPheThrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCys 199
DB 228 GTCAAGCGCAACGAGCGCTGCGGCTGGAAGAGATGCTCGCGGCAATGAAGGGG 169
OY 200 ThrLeuLeuCysProProArgArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsn 219
DB 168 CAGGTGTTTCAAGCCCGCGCGCTTCTGACGACGACGACGACGACGACGACGACGACG 109
OY 220 GlyIleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyr 239
DB 108 GGTGCGCACGCGCTGCGCGCGC-----CAGCATGACGCGCGCGGATGACGGTTC 58
OY 240 GluProLysCysPheLeu 245
DB 57 CAGCGCGCGCTGCGCGATG 40
  
```

## RESULT 9

US-09-252-991A-801

; Sequence 801, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 801  
 ; LENGTH: 1206

; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-801

## Alignment Scores:

| Pred. No.:             | 4.55e-29 | Length:       | 1206 |
|------------------------|----------|---------------|------|
| Score:                 | 291.50   | Matches:      | 87   |
| Percent Similarity:    | 50.41%   | Conservative: | 37   |
| Best Local Similarity: | 35.37%   | Mismatches:   | 99   |
| Query Match:           | 21.05%   | Indels:       | 23   |
|                        |          | Gaps:         | 9    |

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-09-252-991A-801 (1-1206)

```

OY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys---ValGluPheProPheLeu 19
DB 514 ATGGAAGGACACCTGCGCGCCGATGCTGGAAGACGACCGCGGTTCCCGTTCGTC 573
OY 20 ValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPhe 39
DB 574 GCCTTGCTGGTTTCCCGGCTGACACCCAGTGTGCGGGTGAAGGATATCGCGCTAC 633
OY 40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 59
DB 634 CAGTTGCTTGGCAATCGGTGACGATCGCGCGGCAACCTTCGACAAAGCCGCAAG 693
OY 60 ArgLeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyValAlaIleGlu 79
DB 694 CTGATCGGCTG---GGCTATGCC-----GGTGTCCGGAATATGCC 732
OY 80 HisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProPheLysHisAla 99
DB 733 CGCTTGGCGGAGCGCGGCACTCTGCGCGGCTGTGTTCCCGCGGCGAATGCGCATCGC 792
OY 100 LysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLys 119
DB 793 CCGGCGCTGACTTCAAGCTTCAAGCGGCTCAAGACCTTACCTGAAC---ACCTGCGAG 849
OY 120 LysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 139
DB 850 GTTGGCGTGAAGCCCGGCGGACGACGACGACGACGACGACGACGACGACGACGACG 900
OY 140 AlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeu 159
DB 901 CTGGCGTTCACAGCCCGGCGGTGTCAGACCTGCTGATCAAGTCCGCGCGCTTG--- 957
OY 160 PheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGly 179
DB 958 -----AAGCAGACCGGCTG---AAGAAC-----CTGTGATCGCGCGGCGT 996
OY 180 ValAlaSerAspPheThrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCys 199
DB 997 GTCAAGCGCAACGAGCGCTGCGGCTGGAAGAGATGCTCGCGGCAATGAAGGGG 1056
  
```



```

Oy      200 ThrilleauCysProProArArgLeuCysThrAspAsnGlylleMetIleAlaTPan 219
        ::::::::::::::::::::
Db      1057 CAGGGTTCTACGCCCGCCCGGCTTCTTCACCGCAATGCGCGATGACTCACGCC 116
        ::::::::::::::::::::
Oy      220 GlylleGluArgLeuArgAlaGlyLeuGlylleLeuHisAspIleGluGlylleArgTyx 239
        ::::::::::::::::::::
Db      1117 GGCTGCAGCGCGCTGTGCGCCGCG-----CACCATGACGCGCGCGGATCAGCGTC 1167
        ::::::::::::::::::::
Oy      240 GluProlCysProLeu 245
        ::::::::::::::::::::
Db      1168 CAGCGCGCTGCGCGATG 1185
        ::::::::::::::::::::

RESULT 10
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
;               the Haemophilus influenzae Rd Genome, Fragments
;               Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS V6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Alignment Scores:
Pred. No.:          1,68e-23           Length:          1830121
Score:              289.50             Matches:            82
Percent Similarity: 48.70%             Conservative:       30
Best Local Similarity: 35.65%           Mismatches:         95
Query Match:        20.90%             Indels:             23
DB:                  4                 Gaps:                7
US-10-649-273-2_COPY_148_414 (1-267) x US-09-557-884-1 (1-1830121)

Oy      1 MetGluAlaHisAlaLeuThriIleArgLeu--ThraSnlYsValGlupheProPhelEu 19
Db      552881 ATGGAAGGCAATTACTTGCCGCAATGCTTGATGACAATTGACCACGCACTTCTTTGTT 552882
        ::::::::::::::::::::
Oy      20 ValIeuLeuIleSerGlyGlyHisCySLeuLeuAlaLeuValGlnGlyValSerAspPhe 39
        ::::::::::::::::::::

```

```

Db      552821 GCTCGTTGGATCGGAGTGCACACTCAATTAGTCGGTGATGTGATGAGAAATAAT 552762
Qy      40 LeuleuLeuglylvSerleuaspIlealaproGlyaspmcLeuaplysValAlaarg 59
Db      552761 GAAGTAGAGAGAATCTTAATGAATATCCTGCTGCCAAGCCCTTGATTAACAAGAAAA 552702
Qy      60 ArgIeuSerleuIleIleHisProGluCysSerThrmSerGlyGlyAsnAlaileglu 79
Db      552701 TTACTTGGACATA--GATTATCCA-----GGTGGCCGGCACCTTTCT 552653
Qy      80 HisIeuAlalySGInglyAsnArPheHispHeaapIlelvsProProleuHisIsla 99
Db      552662 CGTTTAGCGGAAAAAGTACCGCAAAATCGTTTCACATTTCCACGTCCAATGACAGATCGT 552603
Qy      100 lvsanCySaapPheSerPheThrGlyleuGlnHisvalThrasplysIleIleMetLys 119
Db      552602 GCAGGCCCTTGATTTTTAGTTTTCTTGCTTTTAAAACAATTGCCGCAAAATCAGTTAATCA 552543
Qy      120 LysGIlyvsgIvgIugIyI-----lllegIlyvsgIyGlnlleuSerSerAlaIlaasp 137
Db      552542 GCAATTAAAAAGAGGGCGCAACTGTATGAGGCAA-----ACTAAAGCAGAT 552498
Qy      138 lleaAlaatrIvalGlnHisThrMetAlaCysHisIleuValIysArgThrmIsarGala 157
Db      552497 ATTGCTTATGCTTCCAGATGCGGTGTGATCTCTTCCC----- 552456
Qy      158 lleuPheCySlyvsgIlnArgspIleuProGlnAsnAlaValIleuValAlaser 177
Db      552455 ---ATTAAATGTAG---CGTGCAATTGAAGAAAGAACGCTATAAACGTTAGATGATGGC 552402
Qy      178 GlyGlyValAlaserAspPheTyrlleArgAlaIleuGlnlleuThrzsnAlatNr 197
Db      552401 GGAGGGGAGACGCAATAAAAAACTCCGAGAAACGCTTGGCCACTTAATGC AAAATTTA 552342
Qy      198 GlnCyThrleuLeuCyProProArgIeucYsthrAspAnsglyIleMetIleAla 217
Db      552341 GGTGGCGAAGCTTTTATCTCACTCACTCAATTTTGTACAGATRAITGTCGATGATTGCT 552282
Qy      218 TrpAnsglyIleGlnArgleuArgAlagly 227
Db      552281 TACACAGGTTTTTTACGTTTAAAAACAAGGT 552252

RESULT 11
US-09-643-990A-1/c
; Sequence 1, Application US/09643990A
; Patent No. 6528289
GENERAL INFORMATION:
APPLICANT: Robert D. Fleischmann
Mark D. Adams
Owen White
Hamilton O. Smith
J. Craig Venter
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643, 990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: 08/487,429  
FILING DATE: 1995-06-07  
APPLICATION NUMBER: 08/426,787  
FILING DATE: 1995-04-21  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PB186P1C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-610-5790  
TELEFAX: 310-309-8439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1830121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-643-990A-1  
  
Alignment Scores:  
Pred. No.: 1,686-23 Length: 1830121  
Score: 289.50 Matches: 82  
Percent Similarity: 48.70% Conservative: 30  
Best Local Similarity: 35.65% Mismatches: 95  
Query Match: 20.90% Indels: 23  
DB: 4 Gaps: 7  
  
US-10-649-273-2\_COPY\_148\_414 (1-267) x US-09-643-990A-1 (1-1830121)  
QY 1 MetGUAhAhiAalaleuThriAargLeu---ThrsAnlyVaIgiuPheProPheLeu 19  
DB 552881 ATGGAAAGGCACTTACTTGGCCCAATGCTTGAGACATTCACCGCACTTCTTCTTGT 552822  
QY 20 ValLeuLeuileseGlyGlyhNisCysLeuLeuAlaLeuValGInglyValSerAspPhe 39  
DB 552821 GCTCTGTGGTATCGGGTGGCCACATCAATTAGTGGTGTGATGAGAAATAT 552762  
QY 40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArg 59  
DB 552761 GAAGTATAGCAAACTATATGATGATGCTGCGCCCAAGCTTGTATAAACGCAAAA 552702  
QY 60 ArgLeuSerLeuileLysHisProGlyCysSerThMetSerGlyGlyValAlaIleGlu 79  
DB 552701 TTACTTGAATA---GATTATCCA-----GGTGGCGGGCACTTCT 552663  
QY 80 HisLeuAlaLysGInglyAsnArgPheHisPheAspIleLysProPheLeuHisAla 99  
DB 552662 CGTTTACCGGAAAAGTACGCCAAATCGTTTCACTTCACAGTCCAAATGACAGATCGT 552603  
QY 100 LysAsnCysAspSerPheThrGlyLeuGlnHisValThrAspIleIleMetLys 119  
DB 552602 GCAAGCGCTGATTTAGTTTTCGTTTAAAAAATTCGCGCAATACAGTTAAATCAA 552543  
QY 120 LysGInglyLeuGluGly-----IleGluLysGlyGlnIleLeuSerSerAlaAlaAsp 137  
DB 552542 GCAATTTAAAAACAGAGCGCACTGATGACCA-----ACTAAAGCAGAT 552498  
QY 138 IleAlaIaIaThrValGInHisThrMetAlaCysHisLeuValLysArgThrHisAla 157  
DB 552497 ATGCTTATGCTTCCAAAGTCCGCTGCTGATCTTGGC-----552456  
QY 158 IleLeuPheCysLysGInArgAspLeuProGlnAsnAlaValLeuValAlaSer 177  
DB 552455 ---ATTAAATGTAAG---CGTGCATTGAAGAAACAGGCTATAAAGCTTATGATTCGCG 552402  
QY 178 GlyGlyValAlaSerAspPheTyrlIaArgArgAlaLeuGluIleLeuThrAsnAlaThr 197  
DB 552401 GGAGGGGTGAGCCCAATTAATAAATCCGAGAAACGCTTGGCACTTAATGCAAAATTTA 552342  
QY 198 GInCysThrLeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAla 217  
DB 552341 GGTGGCGAAGTGTATTATCTCAACCTCAATTTTGTACAGATTAAGTGCAGATATGCT 552282

QY 218 TrpAsnGlyIleGluArgLeuArgAlaGly 227  
DB 552281 TACACAGTGTTTTTCACGTTTAAAAACAAGT 552252  
  
RESULT 12  
US-09-902-540-6612  
Sequence 6612, Application US/09902540  
Patent No. 6833447  
GENERAL INFORMATION:  
APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Miegand, Roger C.  
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-10(115849)B  
CURRENT APPLICATION NUMBER: US/09/902,540  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/217,883  
PRIOR FILING DATE: 2000-07-10  
NUMBER OF SEQ ID NOS: 16825  
SEQ ID NO 6612  
LENGTH: 996  
TYPE: DNA  
ORGANISM: Myxococcus xanthus  
US-09-902-540-6612  
  
Alignment Scores:  
Pred. No.: 6,446-28 Length: 996  
Score: 282.00 Matches: 86  
Percent Similarity: 48.93% Conservative: 28  
Best Local Similarity: 36.91% Mismatches: 87  
Query Match: 20.36% Indels: 32  
DB: 4 Gaps: 9  
  
US-10-649-273-2\_COPY\_148\_414 (1-267) x US-09-902-540-6612 (1-996)  
QY 1 MetGUAhAhiAalaleuThriAargLeuThrsAnlyVaI---GluPheProPheLeu 19  
DB 334 CTGAGAGGCACTGCTGCGCATCGGCTTGGAGAGTGGCGCGGACCCGCTTCTT 393  
QY 20 ValLeuLeuileseGlyGlyhNisCysLeuLeuAlaLeuValGInglyValSerAspPhe 39  
DB 394 GGGCTGCTGCTTCCGGGGGACACACAGCCTTACAGAGTGGCAGGCTTACGGGAGTAC 453  
QY 40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArg 59  
DB 454 CGGCTGGTGGCACACCGCGGACGACGCGCGCGGAGGATATGACAAAGCCTGCG 513  
QY 60 ArgLeuSerLeuileLysHisProGlyCysSerThMetSerGlyGlyValAlaIleGlu 79  
DB 514 ATCTCGGCGCTG---CCGATATCCG-----GGTGGCAAGCCCATGCAC 552  
QY 80 HisLeuAlaLysGInglyAsn-----ArgPheHisPheAspIleLysProPro 95  
DB 553 CAGTTGGCGGACAGAGGAAACCGGAGGCCATCGCTTC-----CCGGGC 597  
QY 96 LeuHisAlaLysAsnCysAspSerPheThrGlyLeuGlnHisValThrAspLys 115  
DB 598 GCGGCGCGGCGCAACTTCCAGAGTGTCTTCCGGGTTGAG-----ACG 645  
QY 116 IleIleMetLysLysGInglyLeuGluGlyIleGluLysGlyGlnIleLeuSerSerAla 135  
DB 646 GCGGCTGCGCAACGTCAGAAAGCAGCGTGGCGGAGGCGAGCGCGCTG-----696  
QY 136 AlaAspIleAlaIaIaThrValGInHisThrMetAlaCysHisLeuValLysArgThrHis 155  
DB 697 GCGGATTTGCGGCTCTTCCAGAGGCGCGGAGCGTCTGCTTCCAGAGAG-----750  
QY 156 ArgAlaIleLeuPheCysLysGInArgAspLeuProGlnAsnAlaValLeuVal 175  
DB 751 -----CTGGTGGCGCGCGCGCGCGGCTG-----GCCCAAGAGAGTTGGTG 792

QY 176 AAlaSerGlyValAlaSerAspPheTyrIleArgAlaAlaLeuGluIleLeuThrAsn 195  
DB 793 CTGTGGCGGGCGCTCTCCCGGAACTGCGGCTGCTGTGTGCGGCGGCGGAG 852  
QY 196 AAlaThrGlnCysThrLeuLeuGlyProProArgLeuGlySerThrAspAsnGlyIleMet 215  
DB 853 GAGCGGGGGGTGAACATGTTCTGCGCCCGGTGCGGTGCGGACAGCAATGGCGCATG 912  
QY 216 ILeAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 228  
DB 913 ATTGGCGTGGCGGGGTATGAGGGGTACCGCGGCGCTG 951  
RESULT 13  
US-09-902-540-503/c  
Sequence 503, Application US/09902540  
Patent No. 6833447  
GENERAL INFORMATION:  
APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Wiegand, Roger C.  
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-10(15849)B  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/217,883  
PRIOR FILING DATE: 2000-07-10  
NUMBER OF SEQ ID NOS: 16825  
SEQ ID NO 503  
LENGTH: 2582  
TYPE: DNA  
ORGANISM: Myxococcus xanthus  
US-09-902-540-503

## Alignment Scores:

Pred. No.: 3,15e-27 Length: 2582  
Score: 282.00 Matches: 86  
Percent Similarity: 48.93% Conservative: 28  
Best Local Similarity: 36.91% Mismatches: 87  
Query Match: 20.36% Indels: 32  
DB: 4 Gaps: 9

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-09-902-540-503 (1-2582)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIysVal---GluPheProPheLeu 19  
DB 1195 CTGGAGGGCCACTGCTGGCCATCGGCTGTGGAGGTGGCGCGGAGCGCGCTTCTT 1136  
QY 20 ValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPhe 39  
DB 1135 GGGCTGTGCTTCCGGCGGCGACACAGCTCTACAGAGTGCAGGCTACCGGCACTAC 1076  
QY 40 LeuLeuLeuGlyIysSerLeuAspIleAlaProGlyAspMetLeuAspIysValAlaArg 59  
DB 1075 CGGCTGTGGGCGACGCGCGACGACGCGCGCGGCGGACATATGACAGACGCGTGGC 1016  
QY 60 ArgLeuSerLeuIleIysHisProGluCysSerThrMetSerGlyIysValAlaIleGlu 79  
DB 1015 ATCTCTGGCCCTG---CCGATATCCG---GGTGGGAGCCCATGAC 977  
QY 80 HisLeuAlaIysGlnGlyAsn-----ArgPheHisPheAspIleIysProPro 95  
DB 976 CAGTTGGCCGACAGAGGGGAACCCGAGGCGCATCGCTTC-----CCGCGC 932  
QY 96 LeuHisHisAlaIysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIys 115  
DB 931 GCGCTCCGCGGCAACAATTGACGATGCTCTCTCCGCGTTGAG-----ACG 884  
QY 116 IleIleMetLeuIysGluIysGluGlnGlyIleGluIysGlnIleLeuSerSerAla 135  
DB 883 GCGGTGTCTCACACATGTCAGAGCAACGCGGTGCGCGGAGCGGCGCTG-----833  
QY 136 AlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValIysArgThrHis 155

DB 832 GCGGATTTGGCGGGTCTTTCACAGAGCCGTGCGGACGTGCTGCAAGAG-----779  
QY 156 ArgAlaIleLeuPheCysIysGlnArgAspLeuLeuProGlnAsnAlaValLeuVal 175  
DB 778 -----CTGGGCGCGCGCGCGCGCGCTG-----GCCACAGCACTTGTGT 737  
QY 176 AAlaSerGlyValAlaSerAspPheTyrIleArgAlaAlaLeuGluIleLeuThrAsn 195  
DB 736 CTGTGGCGGGCGCTGCGGAACTGCGGCTGCGGCACTGTGTACAGCGGAGCGGAG 677  
QY 196 AAlaThrGlnCysThrLeuLeuGlyProProArgLeuGlySerThrAspAsnGlyIleMet 215  
DB 676 GAGCGGGGGGTGAACATGTTCTGCGCCCGGTGCGGTGCGGACAGCAATGGCGCATG 617  
QY 216 ILeAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 228  
DB 616 ATTGGCGTGGCGGGGTATGAGGGGTACCGCGGCGCTG 578  
RESULT 14  
US-09-543-681A-2341  
Sequence 2341, Application US/09543681A  
Patent No. 6605709  
GENERAL INFORMATION:  
APPLICANT: GARY BRETTON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
FILE REFERENCE: 2709,1002-001  
CURRENT FILING DATE: 2000-04-05  
PRIOR APPLICATION NUMBER: US 60/128,706  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 8344  
SEQ ID NO 2341  
LENGTH: 1074  
TYPE: DNA  
ORGANISM: Proteus mirabilis  
US-09-543-681A-2341

## Alignment Scores:

Pred. No.: 1,17e-27 Length: 1074  
Score: 280.50 Matches: 81  
Percent Similarity: 46.77% Conservative: 35  
Best Local Similarity: 32.66% Mismatches: 101  
Query Match: 20.25% Indels: 31  
DB: 4 Gaps: 8

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-09-543-681A-2341 (1-1074)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIysVal---GluPheProPheLeu 19  
DB 385 ATGGAGGGCATTTATTAAGCGCGCATGCTGAAGAAAGAACCCAGATTTCTTCTGTG 444  
QY 20 ValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPhe 39  
DB 445 GCGTACTTGTCTCGGGGGGCGCATACAAATTAATGATGAACAGGATGGGGAATAT 504  
QY 40 LeuLeuLeuGlyIysSerLeuAspIleAlaProGlyAspMetLeuAspIysValAlaArg 59  
DB 505 ACCCTGTAGGTGAGTCTATTGATGATGCTGCTGTAAGCATTTGTAATAACAGCCAG 564  
QY 60 ArgLeuSerLeuIleIysHisProGluCysSerThrMetSerGlyIysValAlaIleGlu 79  
DB 565 CTATGGGGCTT---GATTATCC-----GGCGGCGCTGTTTATCA 603  
QY 80 HisLeuAlaIysGlnGlyAsnArgPheHisPheAspIleIysProProLeuHisHisAla 99  
DB 604 AAATGTCACACAGAGTGAAGAGACGTTTCTTTCTGCTCCATGACAGACGA 663  
QY 100 LysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIleIleMetIys 119  
DB 664 CCGGACATCACTTATGTTCTGTGTTAAACCTTGGCGGTAATATGATATGTCGCA 723



GenCore version 5.1.6.  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 16, 2005, 23:18:53 ; Search time 1813.86 Seconds  
(without alignments)  
913.751 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_148\_414

Perfect score: 1385  
Sequence: 1 MEAHALTRLTNKVEPPLV.....DISKEVGEASIKVQPLMEI 267

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 6054689 seqs, 3103772919 residues

Total number of hits satisfying chosen parameters: 12109378

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool\_p/US10649273/runat.15062005.111418.6138/app.query.fasta\_1.1429  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsu62  
-TRANS=human0.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10649273\_@CGN\_1\_1\_1034\_@runat.15062005.111418.6138  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGOQUERY -NEG SCORES=0 -WAIT -DSBLOK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:\*

- 1: /cgn2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/prodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 3: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 4: /cgn2\_6/prodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 6: /cgn2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 7: /cgn2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 8: /cgn2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 10: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 12: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 13: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 14: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 15: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 16: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 17: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 18: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 19: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 20: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 21: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 22: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 23: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 24: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 25: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 26: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length  | DB ID | Description          |
|------------|--------|-------------|---------|-------|----------------------|
| 1          | 1385   | 100.0       | 1416    | 17    | US-10-120-988-177    |
| 2          | 1385   | 100.0       | 1526    | 14    | US-10-067-443-23     |
| 3          | 1385   | 100.0       | 1526    | 18    | US-10-649-273-23     |
| 4          | 1385   | 100.0       | 1526    | 18    | US-10-651-722-23     |
| 5          | 1385   | 100.0       | 2197    | 14    | US-10-067-443-1      |
| 6          | 1385   | 100.0       | 2197    | 18    | US-10-649-273-1      |
| 7          | 1385   | 100.0       | 2197    | 18    | US-10-651-722-1      |
| 8          | 1385   | 100.0       | 2572    | 21    | US-10-480-988-26     |
| 9          | 1358.5 | 98.1        | 1387    | 18    | US-10-649-273-21     |
| 10         | 1358.5 | 98.1        | 1387    | 18    | US-10-649-273-21     |
| 11         | 1358.5 | 98.1        | 1387    | 18    | US-10-651-722-21     |
| 12         | 1358.5 | 98.1        | 1387    | 21    | US-10-887-553A-1047  |
| 13         | 1348   | 97.3        | 1245    | 14    | US-10-012-140-6      |
| 14         | 1348   | 97.3        | 1820    | 14    | US-10-012-140-4      |
| 15         | 1204   | 86.9        | 2208    | 17    | US-10-094-749-400    |
| 16         | 1204   | 86.9        | 2890    | 20    | US-10-723-860-7447   |
| 17         | 824.5  | 59.5        | 14364   | 14    | US-10-067-443-20     |
| 18         | 824.5  | 59.5        | 14364   | 18    | US-10-649-273-20     |
| 19         | 824.5  | 59.5        | 14364   | 18    | US-10-651-722-20     |
| 20         | 468    | 33.8        | 371     | 19    | US-10-430-201-1118   |
| 21         | 468    | 33.8        | 371     | 19    | US-10-430-201-1118   |
| 22         | 422.5  | 30.5        | 1917    | 18    | US-10-424-599-66417  |
| 23         | 359    | 25.9        | 1628    | 19    | US-10-437-963-11249  |
| 24         | 342    | 24.7        | 756     | 14    | US-10-081-051-8      |
| 25         | 342    | 24.7        | 4360    | 14    | US-10-081-051-2      |
| 26         | 321.5  | 23.2        | 1146    | 17    | US-10-282-122A-14674 |
| 27         | 314.5  | 22.7        | 936     | 17    | US-10-282-122A-8315  |
| 28         | 308.5  | 22.3        | 9454    | 17    | US-10-282-122A-26972 |
| 29         | 308.5  | 22.3        | 10750   | 18    | US-10-672-787-38     |
| 30         | 304.5  | 22.0        | 1032    | 17    | US-10-282-122A-31043 |
| 31         | 304.5  | 21.8        | 1000    | 19    | US-10-343-561-50     |
| 32         | 291.5  | 21.0        | 1026    | 9     | US-09-815-242-7701   |
| 33         | 291.5  | 21.0        | 1026    | 17    | US-10-282-122A-30016 |
| 34         | 289.5  | 20.9        | 1029    | 9     | US-09-815-242-6946   |
| 35         | 289.5  | 20.9        | 1029    | 17    | US-10-282-122A-22020 |
| 36         | 289.5  | 20.9        | 1830121 | 17    | US-10-329-670-1      |
| 37         | 289.5  | 20.9        | 1830121 | 20    | US-10-158-865-1      |
| 38         | 285    | 20.6        | 372     | 20    | US-10-357-930-61088  |
| 39         | 280.5  | 20.3        | 1020    | 17    | US-10-282-122A-32254 |
| 40         | 279.5  | 20.2        | 1014    | 9     | US-09-815-242-9682   |
| 41         | 279.5  | 20.2        | 1014    | 17    | US-10-282-122A-39301 |
| 42         | 278.5  | 20.2        | 1014    | 17    | US-10-282-122A-41977 |
| 43         | 277.5  | 20.0        | 1007    | 17    | US-10-282-122A-19320 |
| 44         | 277    | 20.0        | 927     | 17    | US-10-282-122A-25354 |
| 45         | 277    | 20.0        | 1023    | 17    | US-10-282-122A-31809 |

#### ALIGNMENTS

RESULT 1  
US-10-120-988-177  
; Sequence 177, Application US/10120988  
; Publication No. US20030219745A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Goodrich, Ryle  
; APPLICANT: Liu, Chenchua  
; APPLICANT: Ren, Feiyuan  
; APPLICANT: Wang, Duntui  
; APPLICANT: Dmanac, Radoje T.  
; TITLE OF INVENTION: NO. US20030219745A1el Nucleic Acids and  
; FILE REFERENCE: 802CON Polypeptides  
; CURRENT APPLICATION NUMBER: US/10/120,988  
; CURRENT FILING DATE: 2002-04-11

```
; PRIOR APPLICATION NUMBER: 09/774,528
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: pc_fl_genes Version 2.0
; SEQ ID NO 177
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (205)..(1305)
US-10-120-988-177

Alignment Scores:
Pred. No.: 1,18e-169 Length: 1416
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-120-988-177 (1-1416)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB ATGGAGGCTCATGCACTTACTATTAGGTGACCAATTAAGTAGAATTTCTTTTAGTT 561
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB CTTTATATTTCTGGAGGTCACTGTCTGTGGCATTTAGTTCAGAGATTTTCGATTTCGTG 621
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB CTTCTTGGAAAGCTTTGGACATAGCACACGGTGACATGCTTGCAAGGTGGCAAGAGA 681
QY 622 CTTCTTGGAAAGCTTTGGACATAGCACACGGTGACATGCTTGCAAGGTGGCAAGAGA 681
QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 80
DB CTTCTTTTATAATAAATCATCCAGAGTGTCTCCACCATGATGGTGGAAAGCATGAACAT 741
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProPheLeuHisAlaLys 100
DB TTGGCCAAACCAAGAAATATGATTTGATTTGACATCAACCTCCCTTGCAATGCTAA 801
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValAlaThrAspLysIleIleMetLys 120
DB AATTGTGATTTTCTTTACTGACCTTCAACACGTTACTGTAAATAATATATATAAAG 861
QY 121 GluLysGlnGlnGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
DB GAAAGAGAGAAAGATTTGAGAAAGGGCAATCTGTCTTCAGACAGACATTTGCTGCC 921
QY 862 GAAAGAGAGAAAGATTTGAGAAAGGGCAATCTGTCTTCAGACAGACATTTGCTGCC 921
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
DB ACAGTTCAGACACAAATGCGATGTCATCTTGTGAAGAACACATCGGGCTAATCTGTT 981
QY 922 ACAGTTCAGACACAAATGCGATGTCATCTTGTGAAGAACACATCGGGCTAATCTGTT 981
QY 161 CysLysGlnArgAspLeuProGlnAsnAspAlaValLeuValAlaSerGlyVal 180
DB TGTAAACAGAGAACTTTGTAACCTCAAAATATGCAATGATCGTGTGGTGGTGC 1041
QY 982 TGTAAACAGAGAACTTTGTAACCTCAAAATATGCAATGATCGTGTGGTGGTGC 1041
QY 181 AlaSerAspPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
DB GCAAGTACCTTCTATATCCGACAGAGCTCTGGAATTTTAAACAAACGCAACATGCACT 1101
QY 1042 GCAAGTACCTTCTATATCCGACAGAGCTCTGGAATTTTAAACAAACGCAACATGCACT 1101
QY 201 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220
DB TTGTGTGTCTCTCTCCACAGATATGACATGATATGATGATGATGATGATGATGAT 1161
QY 1102 TTGTGTGTCTCTCTCCACAGATATGACATGATATGATGATGATGATGATGATGAT 1161
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGlu 240
DB ATTGAAAGACTAGCTGTGCTGGCATTTTACATGACATAGAGAGCATCCGCTAAGAA 1221
QY 1162 ATTGAAAGACTAGCTGTGCTGGCATTTTACATGACATAGAGAGCATCCGCTAAGAA 1221
QY 241 ProLysCysProLeuGlyValAspIleSerLysGlnGlyValAlaSerGlyVal 260
DB CTTTCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 1161
```

```
DB 1222 CCAAAATGCTCTTTGGAGTAGACATATCAAAAGAGTTGGAGAGCTTCCATAAAAGTA 1281
QY 261 ProGlnLeuLysMetGluIle 267
DB CCAAAATGCTCTTTGGAGTAGACATATCAAAAGAGTTGGAGAGCTTCCATAAAAGTA 1302
DB 1282 CCAAAATGCTCTTTGGAGTAGACATATCAAAAGAGTTGGAGAGCTTCCATAAAAGTA 1281

RESULT 2
US-10-067-443-23
; Sequence 23, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-23

Alignment Scores:
Pred. No.: 1,32e-169 Length: 1526
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-067-443-23 (1-1526)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB ATGGAGGCTCATGCACTTACTATTAGGTGACCAATTAAGTAGAATTTCTTTTAGTT 60
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB CTTTATATTTCTGGAGGTCACTGTCTGTGGCATTTAGTTCAGAGATTTTCGATTTCGTG 120
QY 61 LeuLeuLysSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 60
DB CTTTATATTTCTGGAGGTCACTGTCTGTGGCATTTAGTTCAGAGATTTTCGATTTCGTG 120
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB CTTCTTGGAAAGCTTTGGACATAGCACACGGTGACATGCTTGCAAGGTGGCAAGAGA 180
QY 121 CTTCTTGGAAAGCTTTGGACATAGCACACGGTGACATGCTTGCAAGGTGGCAAGAGA 180
QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 80
DB CTTCTTTTATAATAAATCATCCAGAGTGTCTCCACCATGATGGTGGAAAGCATGAACAT 240
QY 181 CTTCTTTTATAATAAATCATCCAGAGTGTCTCCACCATGATGGTGGAAAGCATGAACAT 240
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProPheLeuHisAlaLys 100
DB TTGGCCAAACCAAGAAATATGATTTGATTTGACATCAACCTCCCTTGCAATGCTAA 300
QY 241 TTGGCCAAACCAAGAAATATGATTTGATTTGACATCAACCTCCCTTGCAATGCTAA 300
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValAlaThrAspLysIleIleMetLys 120
DB AATTGTGATTTTCTTTACTGACCTTCAACACGTTACTGTAAATAATATATAAAG 360
QY 301 AATTGTGATTTTCTTTACTGACCTTCAACACGTTACTGTAAATAATATATAAAG 360
QY 121 GluLysGlnGlnGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
DB GCAAGTACCTTCTATATCCGACAGAGCTCTGGAATTTTAAACAAACGCAACATGCACT 420
QY 361 GCAAGTACCTTCTATATCCGACAGAGCTCTGGAATTTTAAACAAACGCAACATGCACT 420
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
DB ACAGTTCAGACACAAATGCGATGTCATCTTGTGAAGAACACATCGGGCTAATCTGTT 480
QY 421 ACAGTTCAGACACAAATGCGATGTCATCTTGTGAAGAACACATCGGGCTAATCTGTT 480
QY 161 CysLysGlnArgAspLeuProGlnAsnAspAlaValLeuValAlaSerGlyVal 180
DB CTTTCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 1161
```

Db 481 TGTAAAGCAGAGACTGTTGCTCAAAATATGACTGTTGCTGCTGCTGCTC 540  
Qy 181 AAsenAspNhePheThyGlyLeuGlnIleuThraAsnAlaThrGlnCyThr 200  
Db 541 GCAAGTAACCTTCTATATCCGACAGACTCTGGAATTTTAAACAAACGACAGCTGACT 600  
Qy 201 LeuLeuCyseProProProAlaGlyLeuCyThrAspAsnGlyIleMetIleAlaTrpAsnGly 220  
Db 601 TTGTTGTCTCTCTCCAGACTATGACTGATTAATGACATTAATGATGATGATGATGAT 660  
Qy 221 TTTGTTGTCTCTCTCCAGACTATGACTGATTAATGACATTAATGATGATGATGATGAT 240  
Db 661 ATTTGAAAGACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
Qy 241 ProLysCyseProLeuGlyValAspIleSerIleGlyValIleGlyValIleGlyValIle 260  
Db 721 CCAAAATGCTCTCTTGGAGTACATATCAAAAGAACTTGGAGAACTTCCATTAAGTA 780  
Qy 261 ProGlnLeuLysMetGluIle 267  
Db 781 CCACAAATTAATAATGGAGATA 801  
RESULT 3  
US-10-649-273-23  
; Sequence 23, Application US/10649273  
; Publication No. US20040043407A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
; FILE REFERENCE: D0073 CNT  
; CURRENT APPLICATION NUMBER: US/10/649,273  
; PRIOR FILING DATE: 2003-08-27  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 23  
; LENGTH: 1526  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-649-273-23  
Alignment Scores:  
Pred. No.: 1,32e-169 Length: 1526  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 18 Gaps: 0  
US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-649-273-23 (1-1526)  
Qy 1 MetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
Db 1 ATGAGAGCTCATGACTTACTATAGGTTGACCAATAGAGAAATTTCTTTTAGTT 60  
Qy 21 LeuLeuIleSerIleGlyHisCyseLeuAlaLeuValGlnGlyValIleSerAspPheLeu 40  
Db 61 CTTTGAATTTCTGAGAGTCACTGCTGTTGGCATTAATCAAGAGTTTCAAGATTTTCTG 120  
Qy 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValIleArgArg 60  
Db 121 CTTCTTGAAGAGCTTTTGAACATAGCACAGGTACATGCTTGAACAGGTGCGAAGAGA 180  
Qy 61 LeuSerLeuIleLysHisProGluCyseThrMetSerGlyGlyLysAlaIleGlnHis 80  
Db 181 CTTTCTTTAATAAACAATCCAGAGTGTCTCCACCATGATGTGTGGAAGGCATAGAACAT 240  
Qy 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisAlaLys 100

Db 241 TTGGCCAAACAGAGAAATAGATTTCATTTTGAACATCAAACTCCCTTCATCATGCTAA 300  
Qy 101 AsnCyseAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
Db 301 AATTGTGATTTTCTTTTACTGAGCTTCAACAGCTTACTGATTAATAATATGAAAAAG 360  
Qy 121 GluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140  
Db 361 GAAAAAGAGAAAGGATATGAAAGAGGGCAAACTGCTTTCAGACGACAGATGCTGCC 420  
Qy 141 ThrValGlnHisThrMetAlaCyseHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
Db 421 ACAGTACAGACACAAATGGACATGCTATCTTGAAAAGAAACACATGGGCTATTTCTGTT 480  
Qy 161 CysLysGlnArgAspLeuLeuProGlnHisAsnAlaValLeuValAlaSerGlyGlyVal 180  
Db 481 TGTAAAGCAGAGACTGTTTACTCCAAATAATGACATGCTGTTGATCTGCTGCTGCTC 540  
Qy 181 AAsenAspNhePheThyIleArgArgAlaLeuGlnIleLeuThraAsnAlaThrGlnCyThr 200  
Db 541 GCAAGTAACCTTCTATATCCGACAGACTCTGGAATTTTAAACAAACGACACAGTGCAT 600  
Qy 201 LeuLeuCyseProProProAlaGlyLeuCyThrAspAsnGlyIleMetIleAlaTrpAsnGly 220  
Db 601 TTGTTGTCTCTCTCCAGACTATGACTGATTAATGACATTAATGATGATGATGATGAT 660  
Qy 221 TTTGTTGTCTCTCTCCAGACTATGACTGATTAATGACATTAATGATGATGATGATGAT 240  
Db 661 ATTTGAAAGACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
Qy 241 ProLysCyseProLeuGlyValAspIleSerIleGlyValIleGlyValIleGlyValIle 260  
Db 721 CCAAAATGCTCTCTTGGAGTACATATCAAAAGAACTTGGAGAACTTCCATTAAGTA 780  
Qy 261 ProGlnLeuLysMetGluIle 267  
Db 781 CCACAAATTAATAATGGAGATA 801

RESULT 4  
US-10-651-722-23  
; Sequence 23, Application US/10651722  
; Publication No. US20040048302A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
; FILE REFERENCE: D0073 DIV  
; CURRENT APPLICATION NUMBER: US/10/651,722  
; PRIOR FILING DATE: 2003-08-29  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 23  
; LENGTH: 1526  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-651-722-23  
Alignment Scores:  
Pred. No.: 1,32e-169 Length: 1526  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 18 Gaps: 0  
US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-651-722-23 (1-1526)

```

OY 1 MetGluAlaHisAlaLeuThrIleArgLeuHrAsnLysValGluPheProPheLeuVal 20
Db 1 ATGGAGGCTCATGCACTTACTATTAGGTGGACCAATTAAGAATTTCTTTTATGATT 60
OY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db 61 CTTTCATTTCTGGAGGTCACTGTCTGTGGCATTTAGTTCAAGGACTTCCAGATTTTCTG 120
OY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 60
Db 121 CTTCTGGAAGACTTTGACATAGACACAGGTGACATGCTTGACAAAGGTGGCAAGA 180
OY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 80
Db 181 CTTTCTTTAATAAAACATCCAGAGTGTCCACCAAGAGGTGGGAACCCATAGAACAT 240
OY 81 LeuAlaIleGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
Db 241 TTGGCCAAACAGAAATAGATTTCATTTTGACATCAAACTTCCCTTGATCATGTCTAA 300
OY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValIleThrAspLysIleIleMetLysLys 120
Db 301 AATTGTGATTTTCTTTTACTGCACTTCAACACGTTACTGATTAATAATATGAAAAG 360
OY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140
Db 361 GAAAAAGAGAAAGATATTGAGAAAGGGCAAAATCCCTGTTCAGCAGCAACATTTGCTGCC 420
OY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
Db 421 ACGATACAGACACAAATGGCATGTCTTGTGAAAGAACACATGGGCTATTCTGTTT 480
OY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 180
Db 481 TGTAAACAGAGAGACTTTGTAACCTCAAAATATAGCAATCTGTTGCATCTGGTGTGTC 540
OY 181 AlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
Db 541 GCAAGTAACCTTCAATATCCGACAGAGCTCTGGAAATTTTAAACCAACGCAACACAGTCCACT 600
OY 201 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrPasnGly 220
Db 601 TTGTTGTGTCTCTCCACAGCTATGCACTGATTAATGGCATTAATGATTCATAGGAATGCT 660
OY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 240
Db 661 ATTGAAGAGCTACTGCTGCTGGCATTTTACATGACATGAAAGCATCCGCTATGAA 720
OY 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260
Db 721 CCAAAATGTCTCTTGGAGTGAATATCAAAAGAGTTGGAGAGCTTCCATTAAGA 780
OY 261 ProGlnLeuLysMetGluIle 267
Db 781 CCACAAATTAATAATGAGATA 801

RESULT 5
US-10-067-443-1
; Sequence 1, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0

```

```

; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
US-10-067-443-1

Alignment Scores:
Pred. No.: 2,27e-169 Length: 2197
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-067-443-1 (1-2197)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
Db 672 ATGAGGCGCTCAGACACTTACTATTAGGTGACCAATAAAGTAGAATTTCTTTTAACTT 731
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db 732 CTTTGGATTCTTGAGGCGCTGCTGCTGTCATTAGTTCAGAGGCTTCAGATTTCTG 791
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
Db 792 CTTCTGTAAGAGCTCTTGACATGACACACAGTGCACATGCTTGACAGAGTGGCAAGAGA 851
QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGlnHis 80
Db 852 CTTTCTTAATAAACAATCCAGAGTGTCCACCAATGAGTGTGGGAAGCATAGAACAT 911
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
Db 912 TTGGCGAAACAGAGAAATGATTTTCATTTTAAATCAATCAACCTCCCTTGATCATGCTAAA 971
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
Db 972 AATTGTGATTTTCTTTTACTGACCTTAAACACGTTACTGATTAATAATTAATGAAAAG 1031
QY 121 GluLysGlnGlnGlyIleGluLysGlyGlnIleLeuSerSerAlaIleAspIleAlaIa 140
Db 1032 GAAAAGAGGAGGATGATGAGAAGGGCAAAATCTGCTTCAGACGACAGACATTGCTGCC 1091
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysAlaGlnHisArgAlaIleLeuPhe 160
Db 1092 ACAGTACGACGACACAAATGCGATGTCATCTTGTAAMAAACAACATCGGCGATTCTGTT 1151
QY 161 CysLysGlnAspAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
Db 1152 TGTAAGCAGAGAGACTGTGTAACCTCAAAATATGACAGTACGGTTGCACTGGGTGCTC 1211
QY 181 AlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
Db 1212 GCAAGTACTCTTATATTCGCGAGGCTCTGGAAATTTTAAACAAAGCAACACAGTGCAC 1271
QY 201 LeuLeuCysProProProArgLeuLeuCysThrAspAsnGlyIleMetIleAlaIlePheAsnGly 220
Db 1272 TTGTTGTGTCCTCCCTCCACGACTGATGCACTATATATGCAATTATGATGCAATGATGT 1331
QY 221 IleGlnAspGluAspArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleAlaArgTyrGlu 240
Db 1332 ATGTAAGAACACTACGGTGGCTTGCGCAATTTTACATGACATAGAAAGGCACTCCGCTAATA 1391
QY 241 ProLysCysProLeuGlyValAspIleSerLysGlnValGlyGlnAlaSerIleLysVal 260
Db 1392 CCAAAATCTCTCTTGAGTGAATACATACAAAAGAAAGTTGGAGAAAGCTTCCATTAAGAATA 1451
QY 261 ProGlnLeuLysMetGluIle 267

```



Db 1452 CCACAAATTAAAAATGAGATA 1472

RESULT 6

US-10-649-273-1

Sequence 1, Application US/10649273

Publication No. US20040043407A1

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1

FILE REFERENCE: D0073 CNT

CURRENT APPLICATION NUMBER: US/10/649,273

CURRENT FILING DATE: 2003-08-27

PRIOR APPLICATION NUMBER: US 60/266,518

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 10/067,443

PRIOR FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10

NUMBER OF SEQ ID NOS: 71

SOFTWARE: PatentIn version 3.2

SEQ ID NO 1

LENGTH: 2197

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (231)..(1472)

US-10-649-273-1

Alignment Scores:

Pred. No.: 2,27e-169 Length: 2197

Score: 1385.00 Matches: 267

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-649-273-1 (1-2197)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20

DB 672 ATGGAGGCTCATGCACTTATCTATTAGTGTGACCAATTAAGTAGAATTTCTTTTATTGTT 731

QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40

DB 732 CTTTGGATTTCTGGAGGTCACTGCTGTTGGCATTTAGTTCAAGGAGTTTCAGATTTCTG 791

QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 60

DB 792 CTTCTTGGAAAGCTTTTGGACATAGCACAGGTGACATGCTTGCACAGGTGGCAAGAGA 851

QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80

DB 852 CTTTCTTTAATAAATCATTCAGAGTGTCTCCACATAGTGTGGGAAGCCATGAAACAT 911

QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100

DB 912 TTGGCCAAACAAGAAATAGATTTCATTTTGCATCAACTCCCTTGCATCATGCTAAA 971

QY 101 AsnGlyAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120

DB 972 AATTGGATTTTCTTTTACTGACCTTCAACAGTTACTATTAATAATAATAATAAAG 1031

QY 121 GluLysGlnGlyLysIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140

DB 1032 GAAAGAGAGAGAGATTTGAGAGGGGCAATCTGCTTCCAGACGACATGCTGCC 1091

QY 141 ThrValGlnHisThrMetAlaCysHisLysValLysArgThrHisArgAlaIleLeuPhe 160

DB 1092 AAGGTACACACACATGAGATTCATCTTGTAAAGAAACATCGGGCATTCGTGTT 1151

QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 180

Db 1152 TGTAGACAGACAGACTTGTTACTCCAAATATATGACAGTACTGCTTGCATCTGCTGTCTC 1211

QY 181 AlaSerAspPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200

DB 1212 GCAAGTAACTTCTATATCCGAGAGCTGTGAAATTTTAAACAAGCAACAGTGCAT 1271

QY 201 LeuLeuCysProProProArgLeuGlyThrAspAsnGlyIleMetIleAlaThrAsnGly 220

DB 1272 TTGTTGTCTCTCTCCACAGTATGACATGATATATGATATATGATATGATGAAATGCT 1331

QY 221 IleGluArgLeuAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgPheGly 240

DB 1332 ATTAAAGACTACGTCGTGGCATTTTACATGACATAGAAAGGCAATCCGCTATGAA 1391

QY 241 ProLysCysProLeuGlyValAspIleSerLysGlnValGlnGlyAlaSerIleLysVal 260

DB 1392 CCATAATGTCCTCTTGGAGTGAACATATCAAAAGAGTTGAGAGAGCTTCCATATAAGTA 1451

QY 261 ProGlnLeuLysMetGluIle 267

DB 1452 CCACAAATTAAAAATGAGATA 1472

RESULT 7

US-10-651-722-1

Sequence 1, Application US/10651722

Publication No. US20040048302A1

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1

FILE REFERENCE: D0073 DIV

CURRENT APPLICATION NUMBER: US/10/651,722

CURRENT FILING DATE: 2003-08-29

PRIOR APPLICATION NUMBER: US 60/266,518

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 10/067,443

PRIOR FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10

NUMBER OF SEQ ID NOS: 71

SOFTWARE: PatentIn version 3.2

SEQ ID NO 1

LENGTH: 2197

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (231)..(1472)

US-10-651-722-1

Alignment Scores:

Pred. No.: 2,27e-169 Length: 2197

Score: 1385.00 Matches: 267

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-651-722-1 (1-2197)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20

DB 672 ATGGAGGCTCATGCACTTATCTATTAGTGTGACCAATTAAGTAGAATTTCTTTTATTGTT 731

QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40

DB 732 CTTTGGATTTCTGGAGGTCACTGCTGTTGGCATTTAGTTCAAGGAGTTTCAGATTTCTG 791

QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 60

DB 792 CTTCTTGGAAAGCTTTTGGACATAGCACAGGTGACATGCTTGCACAGGTGGCAAGAGA 851

QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80

|    |      |  |      |
|----|------|--|------|
| Db | 852  | CTTTCCTTTAAATAAAACATCCAGAGTGTCCACCATGAGTGGTGGAAAGCATGAAACAT            | 911  |
| Qy | 81   | LeuNalAlayGsiNGiYAsnAArgPhhHi:sphEaspi1:leuYs-Prop:Pol:uHi:Hi:Al:Al:Ys | 100  |
| Db | 912  | TTGGCCAAACAGAGAAATAGATTTTCATTATTTACATCAAACCTCCCTTGCAATCAGCTAA          | 971  |
| Qy | 101  | AsnCYsAspPheSerPheThrGlyLeuGlnHi:sValThraEplysIleIleMetLysYs           | 120  |
| Db | 972  | AATGTGATTTTCTTTTACTGGACTTCAACCGCTACTGATATAATATATATAAAAG                | 1031 |
| Qy | 121  | GIuLySgiuGIuGIYIleGIuLyGIYGIuIleLeuSerSerAlAlAlAspIleAlAl              | 140  |
| Db | 1032 | GAATAAGAGGAAGGTATTTGAGAAAGGGCAATCTGTCTTACAGACGACGATTCGTCC              | 1091 |
| Qy | 141  | ThraValGlnHi:stThrMetAlaCYsHi:sleuValIysArgThrHi:sArgAlAlIleleuPhe     | 160  |
| Db | 1092 | ACAGTACGACACACAAATGCGATTCATCTTTGAAAAAGAACACATCGGCGATTCGTGT             | 1151 |
| Qy | 161  | CysLySgiuArgAspLeuLeuProGlnAsnAsnAlAlAlLeuValAlAlSerGIYGIYAl           | 180  |
| Db | 1152 | TGTATAGCAGAGAGACTGTGTACTCTCAAAATATATGAGATCGGTGTGCAATCGTGAGTGC          | 1211 |
| Qy | 181  | AlAlSerAsnPheYrIleArgArgAlAlAlGlnGluIleLeuThrAsnAlAlThrGlnCYsThr       | 200  |
| Db | 1212 | GCAAGTAACTCTATATCCGACAGAGCTTGAAATTTTAAACAAACGACACAGTGCACT              | 1271 |
| Qy | 201  | LeuLeuCYsProProProArgLeuCYsThraSpsngIYIleMetIleAlAlTrpAsnGIY           | 220  |
| Db | 1272 | TTGTGTGTCTCCTCCACAGACTATGCACTGATATATGCACTTAATGATTCAGTGAATGAT           | 1331 |
| Qy | 221  | IIleGIuArgLeuArgAlAgIYleuGIYrIleleuHi:sAspIleGIuGIYrIleArgTYrGIu       | 240  |
| Db | 1332 | ATTTAAAGACTACGCTGCGGCTTGCGCAATTTTACATGACATAGAAAGGATCCGCGATGAA          | 1391 |

```

US-10-649-273_2_COPY_148_414 (1-267) x US-10-480-988-36 (1-2572)

: PRIOR FILING DATE: 2001-06-22
: PRIOR APPLICATION NUMBER: US 60/303,445
: PRIOR FILING DATE: 2001-07-06
: PRIOR APPLICATION NUMBER: US 60/305,405
: PRIOR FILING DATE: 2001-07-13
: PRIOR APPLICATION NUMBER: US 60/311,442
: PRIOR FILING DATE: 2001-08-09
: PRIOR APPLICATION NUMBER: US 60/314,821
: PRIOR FILING DATE: 2001-08-24
: PRIOR APPLICATION NUMBER: US 60/315,992
: PRIOR FILING DATE: 2001-08-29
: PRIOR APPLICATION NUMBER: US 60/378,205
: PRIOR FILING DATE: 2002-05-03
: NUMBER OF SEQ ID NOS: 56
: SOFTWARE: PERL Program
: SEQ ID NO 36
: LENGTH: 2572
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: OTHER INFORMATION: Incyte ID No: 7632424CB1
US-10-480-988-36

Alignment Scores:
Pred. NO.:
Score: 2.87e-169
Percent Similarity: 1385.00
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 21
Gaps: 0

Length: 2572
Matches: 267
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

```

RESULT 8  
 US-10-480-988-36  
 ; Sequence 36, Application US/10480988  
 ; Publication No. US20050063877A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: GANDHI, Ameena R.; KABLE, Amy E.;  
 ; APPLICANT: SMARNAKAR, Anita; HAPALIA, April J.A.;  
 ; APPLICANT: TRAN, Bao; DUGGAN, Brendan M.;  
 ; APPLICANT: WARREN, Bridget A.; ISON, Craig H.;  
 ; APPLICANT: HONCHELL, Cynthia D.; NGUYEN, Daniel B.;  
 ; APPLICANT: LU, Dyung Alma W.; LEE, Ernestine A.;  
 ; APPLICANT: YUE, Henry; FORSYTHE, Ian J.;  
 ; APPLICANT: BARROSO, Ines; RAMKIMAR, Jayalaxmi;  
 ; APPLICANT: GRIFFIN, Jennifer A.; LI, Joana X.;  
 ; APPLICANT: YANG, Junning; THANGAVELU, Kavitha;  
 ; APPLICANT: GIETZEN, Kimberly J.; DING, Li;  
 ; APPLICANT: BAUGHN, Marian R.; BOROWSKI, Mark L.;  
 ; APPLICANT: YAO, Monique G.; CHAWLA, Nandinder K.;  
 ; APPLICANT: MASON, Patricia M.; GURURAJAN, Rajagopal;  
 ; APPLICANT: LEE, Sally; BECHA, Shanya D.;  
 ; APPLICANT: LEE, Soo Yeun; TRAN, Uyen K.;  
 ; APPLICANT: ELIOTT, Vicki S.; LUO, Wen;  
 ; APPLICANT: SPRAGUE, William W.; TANG, Y. Tom;  
 ; APPLICANT: LU, Yan; ZEBARJADIAN, Yeganeh  
 ; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES  
 ; FILE REFERENCE: PF-1040 USN  
 ; CURRENT APPLICATION NUMBER: US/10/480,988  
 ; CURRENT FILING DATE: 2003-12-16  
 ; PRIOR APPLICATION NUMBER: PCT/US02/19360  
 ; PRIOR FILING DATE: 2002-06-18  
 ; PRIOR APPLICATION NUMBER: US 60/300,508

|    |      |   |      |
|----|------|---|------|
| OY | 1    | MetC1u1u1aH1sa1aLeuThr11AArgLeuThrx1n1ySvA1G1uPheProPheLeuVal | 20   |
| Db | 585  | ATGAGGCTCTATGCACCTTACTATTAGGTGACCAATTAAGATAGATTCTCTTTTAACT    | 644  |
| OY | 21   | LeuLeu1e1eSerG1yG1yH1sCysLeuLeuA1aLeuValG1ng1yValSerAspPheLeu | 40   |
| Db | 645  | CTTTTGATTCTTGAGGGTCACTGTCGTGTGGCATTTAGTTCAAAGAGTTTCAGATTCTG   | 704  |
| OY | 41   | LeuLeuG1yLysSerLeuAsp11eAlaProG1yAspMetLeuAspLysValAlaArgArg  | 60   |
| Db | 705  | CTTCTTGAAAGTCTTTGGACATGACACACAGTATACGCTTGGCAAGGTGGCAAGAGA     | 764  |
| OY | 61   | LeuSerLeu11eLysSH1sProG1yCysSerThrMetSerG1yG1yLysAla11eG1uH1s | 80   |
| Db | 765  | CTTCTTTTAATTAANAACATCCAGAGTGCTCCACCATGAGTGGTGGGAAGCCATAGACAT  | 824  |
| OY | 81   | LeuAlaLysG1ng1yAsnArgPheH1sPheAsp11eLysProProLeuH1sH1sAlaLys  | 100  |
| Db | 825  | TTGGCCAAACAGAGAANAATGATTTCATTTTGACATCAAACTCCCTTGSCATCATGCTAA  | 884  |
| OY | 101  | AsnCysAspPheSerPheThrG1yLeuG1nH1sValThraAspLys11e11eMetLysLys | 120  |
| Db | 885  | AATGTGATTTTTCTTTTACTGACCTTCACACCTTACTGATTAANAATTAATATGAANAAG  | 944  |
| OY | 121  | G1uLysG1uG1uG1yL1eG1uLysG1yG1n11eLeuSerSerAla1aAsp11eAla1a    | 140  |
| Db | 945  | GAANAAGAGGAGGATTTGAGAGGGGGCAAACTCTGTTTACGACGACAGCATTCCTGCC    | 1004 |
| OY | 141  | ThrValG1nH1sThrMetAlaCysH1sLeuValLysArgThrH1sArgAla11eLeuPhe  | 160  |
| Db | 1005 | ACAATTAACGACACACATATGCATGTCATCTTGGAAGAAACACATCGGGCATTTCTGTTT  | 1064 |
| OY | 161  | CysLysG1nArgAspLeuLeuProG1nAsnAsnAlaValLeuValAlaSerG1yG1yVal  | 180  |
| Db | 1065 | TGTAAGCAGAGAAGCTTGTTACCTCAAAATATATCAGATACGTGGTCATCTGGTGCTC    | 1124 |
| OY | 181  | AlaSerAsnPheTyrr11eArgArgAlaLeuG1u11eLeuThraAsnAlaThG1CysThr  | 200  |
| Db | 1125 | GCAAGTAACTTCTATATCTCCGAGAGCTCTGGAATAATTTTAACAACGCAACAAGTGACT  | 1184 |

Oy 201 LeuLeuCyseProProPcoAcyLeuCyseThraspaNGly11MeMet11ealTTPaNGly 220  
 Db 1185 TTTGTGTGTCTCTCCAGACTATGCACTATATATGGCATATATGATTCAGAGAAATGGT 1244  
 Oy 221 11eGuArgLeuArgAlaGlyLeuGly11Leuuh1saP11eGuGly11eArgTyrGlu 240  
 Db 1245 ATGTAAAGACTAGTGTGTGGCTTGGCATTTTACATGACATGAAGAGCATCGCTATGA 1304  
 Oy 241 ProLyCyseProLeuGlyValAsp11eSerLyGluValGlyGluAlaSer11eLyVal 260  
 Db 1305 CCAAAATGTCTCTTGGAGTAGACATATCAAAAGAAATGGAGAACTCCATAAAAAGTA 1364  
 Oy 261 ProGlnLeuLyMetGlu11e 267  
 Db 1365 CCACATTAATAAAATGGAGATA 1385  
 RESULT 9  
 US-10-067-443-21  
 ; Sequence 21, Application US/10067443  
 ; Publication No. US20030082782A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bristol-Myers Squibb Company  
 ; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
 ; TITLE OF INVENTION: SPINAL CORD, MP-1  
 ; FILE REFERENCE: D0073 NP  
 ; CURRENT APPLICATION NUMBER: US/10/067,443  
 ; CURRENT FILING DATE: 2002-02-05  
 ; PRIOR APPLICATION NUMBER: US 60/266,518  
 ; PRIOR FILING DATE: 2001-02-05  
 ; PRIOR APPLICATION NUMBER: US 60/282,814  
 ; PRIOR FILING DATE: 2001-04-10  
 ; NUMBER OF SEQ ID NOS: 71  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 21  
 ; LENGTH: 1387  
 ; TYPE: DNA  
 ; ORGANISM: homo sapiens  
 US-10-067-443-21  
 Alignment Scores:  
 Pred. No.: 3,25e-166 Length: 1387  
 Score: 1358.50 Matches: 266  
 Percent Similarity: 91.10% Conservative: 0  
 Best Local Similarity: 91.10% Mismatches: 1  
 Query Match: 98.09% Indels: 25  
 DB: 14 Gaps: 1  
 US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-067-443-21 (1-1387)  
 Oy 1 MetGluAlaHisAlaLeuThr11eArgLeuThrAsnLySValGluPheProPheLeuVal 20  
 Db 465 ATGGAGGCTCATGCACTTACTATTAGTGTGACCAATAAAGTGAATTTCTTTTAAAGTT 524  
 Oy 21 LeuLeu11eSerGlyGlyHisCyseLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
 Db 525 CTTTATATTTCTGAGAGGTCACGTGCTGTGTGGCATTAATTCAGAGCATTTTCATG 584  
 Oy 41 LeuLeuGlyLySserLeuAsp11eAlaProGlyAspMetLeuAspLySValAlaArg 60  
 Db 585 CTTCTTGAAGAAGCTTTTGGACATAGCACCAAGTGACATCTTGCACAAAGTGGCAGAGAGA 644  
 Oy 61 LeuSerLeu11eLySHisProGlnCyseSerThrMetSerGlyGlyLySAla11eGluHis 80  
 Db 645 CTTTCTTATTAATAAACATCCAGAGTGTCCACCAAGAGTGGGGAAGGCATAGAGCAT 704  
 Oy 81 LeuAlaLySglnGlyAsnArgPheHisPheAsp11eLySProProLeuHisHisAlaLyS 100  
 Db 705 TTGGCCAAACAAGAAATAGATTTCATTTTGCATCAAACTCCCTTGCATCATGCTAAA 764  
 Oy 101 AsnCyseAspPheSerPheThrGlyLeuGlnHisValThrAspLyS11eMetLyLyS 120  
 Db 765 AATTTGATTTTCTTTTACTGACCTTCAACAGTTAAGTAATTAATTAATGAAGAAAG 824

|  |      |  |      |
|--|------|--|------|
| Oy   | 121  | GIUlyeGIuGIuGIyIle-----  | 156  |
| Db   | 825  | GAABAAGAGAAAGATATATTTCTAATTAAGTTGAACAGATAATATCTCGATTG          | 884  |
| Oy   | 127  | -----GIUlyeGIyGIuIleSeuSerAla                                  | 135  |
| Db   | 885  | TGCCTAAATAATAGCTCATTTCTGCAGGTATGAGAAAGGGCAATCTGCTTCAGCA        | 944  |
| Oy   | 136  | AlAaSpIleAlaIeThVaIGInHieThMeAlAcyehIleuVallysaATThHis         | 155  |
| Db   | 945  | GCAGACATTTGCTGCCACAGTACAGACACCAATGGCATTCATTGTGAAAAGAACACAT     | 100  |
| Oy   | 156  | ArgAlaIleleuPheCyAlaIeSgInHrAspLeuProGlnAsnAsnAlaValleuVal     | 175  |
| Db   | 1005 | CGGGCTATTCCTGTTTGTGPAAGAGAGAGACTTGTAACCTCAAAATATATGACGTACTGGIT | 106  |
| Oy   | 176  | AlAserGIyGIyValAlAserAsnPherYrIleArgARAlaleuGIuIleuThrAsn      | 199  |
| Db   | 1065 | GCATTTGATGCTGTGCGCAAGTACTTCTATATCCGAGAGCTTGGAATTTTAAACAAC      | 112  |
| Oy   | 196  | AlaThrgInCySthrLeuLeuCyAsProProPoaArgLeuCySthraSpsaNGIyIleMet  | 215  |
| Db   | 1125 | GCAACACAGTGACCTTTGTGTGTCTCTCCAGACTATGACGTATATGACATATATG        | 118  |
| Oy   | 216  | IleAlaTPaNGIyIleGIuArGLeulArGAlaGIyLeuGIyIleleuHIsaSpIleGIu    | 235  |
| Db   | 1185 | ATTCGATGAATGATGATTAAGAAAGACTACCTGGTGGCTTTTACATGACATAGAA        | 124  |
| Oy   | 236  | GIyIleArGTYrGIuProLySeCySProLeuGIyValAspIleSeerlySGIyValGIyGu  | 255  |
| Db   | 1245 | GGCATCCCGCTTGAACCAAAAGTCTCTTGGAGTACACTATCAAAAGAAAGTTGAGAA      | 130  |
| Oy   | 256  | AlAserIleLySValProGInLeuLySMeGIuIle                            | 267  |
| Db   | 1305 | GCTTCATTAAGTACCAACATTTAAATATGAGATA                             | 1340 |
| RESULT 10  |      |  |      |
| US-10-649-273-21   |      |  |      |
| Sequence 21, Application US/10649273                                     |      |  |      |
| Publication No. US20040043407A1  |      |  |      |
| GENERAL INFORMATION:   |      |  |      |
| APPLICANT: Bristol-Myers Squibb Company                                  |      |  |      |
| TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1 |      |  |      |
| FILE REFERENCE: D0073 CNT  |      |  |      |
| CURRENT APPLICATION NUMBER: US/10/649,273                                |      |  |      |
| CURRENT FILING DATE: 2003-08-27  |      |  |      |
| PRIOR APPLICATION NUMBER: US 60/266,518                                  |      |  |      |
| PRIOR FILING DATE: 2001-02-05  |      |  |      |
| PRIOR APPLICATION NUMBER: US 10/067,443                                  |      |  |      |
| PRIOR FILING DATE: 2002-02-05  |      |  |      |
| PRIOR APPLICATION NUMBER: US 60/282,814                                  |      |  |      |
| PRIOR FILING DATE: 2001-04-10  |      |  |      |
| NUMBER OF SEQ ID NOS: 71   |      |  |      |
| SOFTWARE: PatentIn version 3.2   |      |  |      |
| SEQ ID NO 21   |      |  |      |
| LENGTH: 1387   |      |  |      |
| TYPE: DNA  |      |  |      |
| ORGANISM: homo sapiens   |      |  |      |
| US-10-649-273-21   |      |  |      |
| Alignment Scores:  |      |  |      |
| Pred. No.: 3,25e-166 Length: 1387  |      |  |      |
| Score: 1358.50 Matches: 266  |      |  |      |
| Percent Similarity: 91.10% Conservative: 0                               |      |  |      |
| Best Local Similarity: 91.10% Mismatches: 1                              |      |  |      |
| Query Match: 98.09% Indels: 25   |      |  |      |
| DB: 18 Gaps: 1   |      |  |      |
| US-10-649-273-2_COPY_148_414 (1-267) x US-10-649-273-21 (1-1387)         |      |  |      |
| Oy   | 1    | MetGIuAlaHIAIAlaIeThrIleArGLeulThrAsnLySValGIuPheProPheLeuVal  | 20   |
|  |      |  |      |

|   |      |   |     |
|---|------|---|-----|
| Db  | 465  | ATGAGGCTCATGCACTACTACTATTATGGTTGACCAATAAAGTAGAATTTCTTTTAGTT     | 522 |
| Qy  | 21   | LeuLeuLleSerGlyGlyYhiSeCySeLeuLeuAlaLeuValGlnGlyValSerAspPheLeu | 40  |
| Db  | 525  | CTTTTGATTTCTGAGGTCACGTCTGTGGACATAGTTCAAGAGACTTCAGATTTTCTG       | 584 |
| Qy  | 41   | LeuLeuGlyLysSerLeuAspLleAlaProGlyAspMetLeuAspLysValAlaArg       | 60  |
| Db  | 585  | CTTCTTGAAAGTCTTTTGACATAGCACAGGTGACATCTTGACAAAGGTGGCAAGA         | 644 |
| Qy  | 61   | LeuSerLeuLleYshIseProGluCySerThrMetSerGlyGlyYsaLleGlnHis        | 80  |
| Db  | 645  | CTTTCTTTAATPAAACATCCAGAGTGCTCCACATGAGTGGTGGAAACCATAGAGCAT       | 704 |
| Qy  | 81   | LeuAlaLysGlnGlyAsnArgPheHisPheAspLleYseProProLeuHisshisAlaYs    | 100 |
| Db  | 705  | TTGGCCAAACAAAGAAATAGATTTCAATTTGACATCAACCTCCCTTGACATCTTAA        | 764 |
| Qy  | 101  | AsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspLysLleLleMetLysLys     | 120 |
| Db  | 765  | AATTGTGATTTTCTTTTACTGACTTCAACAGCTTACTGATTAATAATTAATGAAG         | 824 |
| Qy  | 121  | GluYsgGlnGlyYlle-----GluYsgGlyGlnLleLeuSerSeraIa                | 126 |
| Db  | 825  | GAAAAAGAGGAAGATATTTCTTAATAGTAAGTTGAACAGATAAATTTCTTGATTG         | 884 |
| Qy  | 127  | -----GluYsgGlyGlnLleLeuSerSeraIa                                | 135 |
| Db  | 885  | TGCGCTAAATAATAGCTCATTTCTTGACAGTATGAGAAAGGGCAAAATCTGTCTTCAG      | 944 |
| Qy  | 136  | AlaAspLleAlaAlaThrValGlnHisThrMetAlaCyshisLeuValLysArgThrHis    | 155 |
| Db  | 945  | GCAAGCATTTGCTGCCACAGTACAGACACAATGCGATGTCATCTGTGTGAAGAAACAT      | 100 |
| Qy  | 156  | ArgAlaLleLeuPheCySLeuArgAspLeuLeuProGlnAsnAsnAlaValLeuVal       | 175 |
| Db  | 1005 | CGGCGTATTTCTTTTGTATAGCAGAGAGCTTTTACCTCAAAATATATGACATCGGTT       | 106 |
| Qy  | 176  | AlaSerGlyGlyValAlaSerAsnPheYrLleArgArgAlaLeuGlnLleLeuThrAsn     | 195 |
| Db  | 1065 | GCATCTGGTGGTGGCAAGTAACTTCTATATCCGACAGAGCTCTGGAAATTTTAAACAAC     | 112 |
| Qy  | 196  | AlaThrGlnCySThrLeuLeuCySProProProArgLeuCySThrAspAsnGlyLleMet    | 215 |
| Db  | 1125 | GCAACACAGTGACATTTGTGTCTCTCTCCACAGCTATGACATGATATGACATTATG        | 118 |
| Qy  | 216  | LleAlaThrAsnGlyLleGluArgLeuArgAlaGlyLeuGlyLleLeuHisAspLleGln    | 235 |
| Db  | 1185 | ATTGCATGAAATGGATTTGAAAGACTACGTGGCTTGAGCAATTTTACATGACATAGAA      | 124 |
| Qy  | 236  | GlyLleArgYrGlnProLysCySProLeuGlyValAspLleSerLysGlnValGlyGln     | 255 |
| Db  | 1245 | GGCATCCGCTATGAAACCAAAATGCTCTCTTGAGATAGACATATCAAAAGAAGTTGAGA     | 130 |
| Qy  | 256  | AlaSerLleLysValProGlnLeuLysMetGluLle 267                        |     |
| Db  | 1305 | GCTTCCTCAATAAAGTACCAAAATTAATAATGAGATA 1340                      |     |
| RESULT 11   |      |   |     |
| US-10-651-722-21  |      |   |     |
| Sequence 21, Application US/10651722                                      |      |   |     |
| Publication No. US20040048302A1   |      |   |     |
| GENERAL INFORMATION:  |      |   |     |
| APPLICANT: Bristol-Myers Squibb Company                                   |      |   |     |
| TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1 |      |   |     |
| FILE REFERENCE: D0073 DIV   |      |   |     |
| CURRENT APPLICATION NUMBER: US/10/651,722                                 |      |   |     |
| CURRENT FILING DATE: 2003-08-29   |      |   |     |
| PRIOR APPLICATION NUMBER: US 60/266,518                                   |      |   |     |
| PRIOR FILING DATE: 2001-02-05   |      |   |     |
| PRIOR APPLICATION NUMBER: US 10/067,443                                   |      |   |     |
| PRIOR FILING DATE: 2002-02-05   |      |   |     |
| PRIOR APPLICATION NUMBER: US 60/282,814                                   |      |   |     |

```

; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 21
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-651-722-21

Alignment Scores:
Pred. NO.:      3,25e-166      Length:      1387
Score:          1358.50       Matches:      266
Percent Similarity: 91.10%   Conservative: 0
Best Local Similarity: 91.10% Mismatches:     1
Query Match:    98.09%       Indels:      25
DB:             18           Gaps:        1

US-10-649-273-2_COPY_148_414 (1-267) x US-10-651-722-21 (1-1387)

QY      1 MetGUAIAHISALaleuthrTlIeaRgleuthrTAenlySvalGIuPherProPheLeuVal 20
Db      465 ATGGAGGCTCATTGCACCTTACTATTTAGGTGAACAATAAAGTAGAAATTCCTTTTTAGTT 524

QY      21 LeuleuileSerGIyGIyHIseCySeuleuualaleuValIGnGIyValseAspPheleu 40
Db      525 CTTTGATTTCTGGAGTCACTGCTGTGGCATTTAGTTCAAGGACTTTCAGATTTTCTG 584

QY      41 LeuleugIyLSerSeLeuaspIIleaIaprogIyaSpMetLeuaspLySvalAlaarArg 60
Db      585 CTTCCTGGAAAAGTCTTTGGACATACACCAAGGTGACATGCTTGACAAAGGTGCCAAGA 644

QY      61 LeuserleuIIelySHisPprogIucYserThrMetSerGIyGIyLysalIlegIuHis 80
Db      645 CTTTCTTTAAATAAACATCCAGAGTGTCCAACATGAGGTGGGGAACCATAAGAGCAT 704

QY      81 LeuulalySGnGIyASnarIPhehisPheaSPIleySProProleuhISHIsalLyS 100
Db      705 TTGGCGAAACAAAGAAATGATTTATTTGACATCAAACCTCCCTTGCAATAGCTAA 764

QY      101 AsnCYasPheSerPheThrGIyLeuGINHISvalThrasPIySIleIemEctLYS 120
Db      765 AATTGTGATTTTCTTTTACTGGAATTCAACGTTACTGATTAATAATATGAAGAAAG 824

QY      121 GIuLySGluGIyLIle----- 126
Db      825 GAAAAAGAGAGAGATATTTTCTAATTAAGTAAGTTGAACAGATATAATATTCCTGGATTG 884

QY      127 -----GIuLySGIyGnIleuserSerAla 135
Db      885 TGCTTAAATAATAGCTGCTCATTTCTGCAGGTATGAGAAAGGGCAAAATCCTGCTTACGA 944

QY      136 AlaaSPILleaIaaIathrValGINHISThrmecAlaCYSHISLeuValIySarGTNHS 155
Db      945 GCAGACATTTGCTGCGACAGTACAGACACAAATGCGATGTCACTTCTGAAAGAACAT 100

QY      156 ArgAlaIleleuPheCySLySGlnarGaspleuLeuProGlnasnAnaIavaIleuVal 175
Db      1005 CGGGCTATTCTGTGTTTGTATAGCAGAGACCTGTTAACCTCAAAATATATCACAGTACGTG 106

QY      176 AlaSerGIyGIyValAlaserAsmPheYrIIleaRGARAlaLeuGluIIleuThraSn 195
Db      1065 GCATCTGGGTGTGTCGCAAGTAACTTCATATACCGCAGAGCTCTGGAATTTTAAACAAC 112

QY      196 AlaThrGInCYSThrLeuLeuCYSProProAArgLeuCYSThrAspAsnGIyIleMet 215
Db      1125 GCAACACAGTGCATCTTTGTGTGTCCTCTCCCAACATATGCACTGATATATGGCATTTATG 118

QY      216 IleaIatPaSngIyIleGIuArGPeuaRgaIagIyleuGIyIleuNHISaPIlegIu 235
Db      1185 ATTGATATGAATGGATATGAAGAAGTACGTGTGGCTTGAGCATTTTATCATGACATAGAA 124

QY      236 GIyIleaRgIyRGIuProLyCySProLeuGIyValaSPIleserIySGIuValIGIyGIu 255

```

Db 1245 GGATCCGCTATGTAACCAAAATGCTCTTGAGTACATATATCAAAAGATTGGAGAA 1304  
Qy 256 ALaSerIleYsValProGlnLeuYsMetGluIle 267  
Db 1305 GCTTCCATTAAGTACCAACATTTAAAAATGGAGATA 1340  
RESULT 12  
US-10-887-553A-1047  
; Sequence 1047, Application US/10887553A  
; Publication No. US20050085436A1  
; GENERAL INFORMATION:  
; APPLICANT: Garza, Dan  
; APPLICANT: Li, Hao  
; TITLE OF INVENTION: Method to treat conditions associated  
; TITLE OF INVENTION: with insulin signalling dysregulation  
; FILE REFERENCE: 4-33262  
; CURRENT APPLICATION NUMBER: US/10/887, 553A  
; PRIOR FILING DATE: 2004-07-08  
; PRIOR APPLICATION NUMBER: 60/485,883  
; NUMBER OF SEQ ID NOS: 1208  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1047  
; LENGTH: 1387  
; TYPE: DNA  
; ORGANISM: human  
US-10-887-553A-1047  
Alignment Scores:  
Pred. No.: 3,25e-166 Length: 1387  
Score: 1358.50 Matches: 266  
Percent Similarity: 91.10% Conservative: 0  
Best Local Similarity: 91.10% Mismatches: 1  
Query Match: 98.09% Indels: 25  
DB: 21 Gaps: 1  
US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-887-553A-1047 (1-1387)  
Qy 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnYsValGluPheProPheLeuVal 20  
Db 465 ATGAGAGCTCATGACCTACTACTATTAGTTGACCAATAAGTAGAATTTCTTTTAACTT 524  
Qy 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
Db 525 CTTTGATTTCTGAGAGGCTGCTGCTGTTGCAATTAGTTCAAGGATTTCAGATTTCTG 584  
Qy 41 LeuLeuGlyYsSerLeuAspIleAlaProGlyAspMetLeuAspYsValAlaArgArg 60  
Db 585 CTTCTTGAAAGTCTTTGGACATAGCACACAGTGACATGCTTGACAAAGTGGCAAGAGA 644  
Qy 61 LeuSerLeuIleYsHisAspGluCysSerThrMetSerGlyGlyValAlaIleGluHis 80  
Db 645 CTTTCTTTAATRAAACAATCCAGAGTCTCCACATGAGTGGGGAAGCCATAGAGCAT 704  
Qy 81 LeuAlaIleGlnGlyAsnArgPheHisAspAspIleYsProProLeuHisHisAlaYs 100  
Db 705 TTGGCCCAACAAAGAAATAGATTCTATTGACATCAAACTCCCTTCATCATCTAA 764  
Qy 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspYsIleIleMetYsYs 120  
Db 765 AATTGTGATTTCTTTTACTGAGCTTCAACAGCTTACTGATTAATAATATGAAAAAG 824  
Qy 121 GluYsGluGluGlyIle-----GluYsGlyGlnIleLeuSerSerAla 135  
Db 825 GAAAAAGAGGATATATTTCTAATTAGTAAGTTGAACAGATAATATTCCTGATTC 884  
Qy 127 -----GluYsGlyGlnIleLeuSerSerAla 135  
Db 885 TGCTTAATAATAGCTCATTTCTGACAGTATGAGAGGGCAATCTCTGCTTCAGCA 944  
Qy 136 AlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeuValYsArgArgHis 155  
Db 945 GCAGACATTCGTCACAGTACAGACACAAATGGCATGATCTGTTGAAAAAGACACAT 1004

Qy 156 ArgAlaIleLeuPheCysYsGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuVal 175  
Db 1005 CGGGCTATTCTGTTTGTAAAGACAGACATTTTACTCCAAAAATATGCACTGCTT 1064  
Qy 176 ALaSerGlyYsValAlaSerAsnPheYrIleArgArgAlaLeuGlnIleLeuThrAsn 195  
Db 1065 GCATCTGTGTGTGCGAAGTACTCTATATCCAGAGCTCTGAAATTTTAACAAAC 1124  
Qy 196 AlaThrGlnCysThrLeuLeuCysProProProAlaGlyLeuCysThrAspAsnGlyIleMet 215  
Db 1125 GCACACAGTGACCTTTGTGTGCTCTCTCCACACTATGACATGATATGCACTTANG 1184  
Qy 216 ILeaIATPaSngIYlIeGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlu 235  
Db 1185 ATTGATGGAATGATTAAGAAGACCTACGTTGCTGGGCACTTTTACATGACATAGA 1244  
Qy 236 GlyIleArgYrGluProYsCysProLeuGlyYsAlaAspIleSerYsGluValGlyGlu 255  
Db 1245 GGATCCGCTATGAAACCAAAATGCTCTTGAGTAGACATATCAAAAGAGTTGGAGAA 1304  
Qy 256 ALaSerIleYsValProGlnLeuYsMetGluIle 267  
Db 1305 GCTTCCATTAAGTACCAACATTTAAAAATGGAGATA 1340  
RESULT 13  
US-10-012-140-6  
; Sequence 6, Application US/10012140  
; Publication No. US20030009017A1  
; GENERAL INFORMATION:  
; APPLICANT: Leiby, Kevin R.  
; APPLICANT: Kapeller-Livermann, Rosana  
; APPLICANT: Glucksmann, Maria A.  
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND  
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES  
; FILE REFERENCE: 381552004900  
; CURRENT APPLICATION NUMBER: US/10/012,140  
; PRIOR FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: 60/246,768  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/246,772  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/249,185  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 1245  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-012-140-6  
Alignment Scores:  
Pred. No.: 6,47e-165 Length: 1245  
Score: 1348.00 Matches: 260  
Percent Similarity: 98.50% Conservative: 3  
Best Local Similarity: 97.38% Mismatches: 4  
Query Match: 97.33% Indels: 0  
DB: 14 Gaps: 0  
US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-012-140-6 (1-1245)  
Qy 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnYsValGluPheProPheLeuVal 20  
Db 442 ATGAGAGCTCATGACCTACTACTATTAGTTGACCAATAAGTAGAATTTCTTTTAACTT 501  
Qy 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
Db 502 CTTTGATTTCTGAGAGTCACTGCTGTTGCAATTAGTTCAAGGAGTTTCAGATTTTCTG 561  
Qy 41 LeuLeuGlyYsSerLeuAspIleAlaProGlyAspMetLeuAspYsValAlaArgArg 60

Db 562 CTTCTTGGAAGCTTTGGACATAGCAGCCAGTGACATGCTTGCACAAAGTGGCAAGAGA 621  
Qy 61 LeuSerLeuIlelyshisProgluCySerThrmeserlyglylysaalilegluhis 80  
Db 622 CTTCTTTTATAAATCAATCCAGAGTGCTCCACCATGATGTGGGAAGCATAGAACAT 681  
Qy 81 LeuAlaIyGlnGlysnatgphelaphaspilleyserproleuhsiallys 100  
Db 682 TTGGCCCAACAGAAATATGATTCATTTGACATCAACCTCCCTTGACATCATGCTAAA 741  
Qy 101 AasnCyAspPheSerPheThrglyleuglnhisvalthrapslylleilemetlylys 120  
Db 742 AATGTGATTTTCTTTTACTGAGCTTCAACAGCTTCTATTAATAAATGAAACAGG 801  
Qy 121 GlnyGlygluglylileglulysglylnlleuSerSerAlaIleaspilleala 140  
Db 802 AAACAAGAGGAAGTATTGAGAGGGGCAATCTGTCTTCAGAGAGACATTTGCTGCC 861  
Qy 141 ThrValGlnhisThrmelAlaCyshisLeuVallysaThrhIsarGAlaIleleuphe 160  
Db 862 ACAGTACAGACACACAATGGCATGTCATCTGTGAAAGAACACATCGGCTATTCTGTTT 921  
Qy 161 CysIyGlnArgAspLeuProglInasnAsnAlaValleuValAlaSerGlyVal 180  
Db 922 TGTAAAGCAGAGAGACTGTGATCTCAAAATATGACAGTACTGTGATCTGTGATGTC 981  
Qy 181 AlaSerAspPheThrhIleargArgAlaLeuGlnlleuThrhAsnAlaThrhInCysThr 200  
Db 982 GCAAGTAACTTCTATATCCGACAGCTCTGGAAATTTTAAACAAGACACAGTCACT 1041  
Qy 201 LeuLeuCySerProProArgLeuCySerThrhAspAsnGlyllemetllealatrpsnGly 220  
Db 1042 TTGTTGTGTCCTCTCCACAGACTATGACATGATATGATGATGATGATGATGATGAT 1101  
Qy 221 IleGlnArgLeuArgAlaGlyleuglylleleuhsiaspillegluglylleargTyGlu 240  
Db 1102 ATTTGAAAGACTAGCTGCTGCTGGCATTTTACATACATAGAGGACATCCGCTATGAA 1161  
Qy 241 ProLyCySProlLeuGlyValAspIleSerLyGlnValGlylnAlaSerIlelyVal 260  
Db 1162 CCAAAATGTCCTTTGAGTAGACATATCAAAAGAGTGGAGAGCTTCATTAAGATA 1221  
Qy 261 ProGlnLeuIyMetGluIle 267  
Db 1222 CCACATTAATAATGGAGATA 1242

RESULT 14  
US-10-012-140-4  
; Sequence 4, Application US/10012140  
; Publication No. US2003009017A1  
; GENERAL INFORMATION:  
; APPLICANT: Leiby, Kevin R.  
; APPLICANT: Kapeller-Libermann, Rosana  
; APPLICANT: Glucksmann, Maria A.  
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND  
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES  
; FILE OF INVENTION: THEREOF  
; PILE REFERENCE: 381552004900  
; CURRENT APPLICATION NUMBER: US/10/012,140  
; CURRENT FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: 60/246,768  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/246,772  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/249,185  
; PRIOR FILING DATE: 2000-11-15  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 1820  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:

NAME/KEY: CDS  
; LOCATION: (146) ... (1390)  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1) ... (1820)  
; OTHER INFORMATION: n = A, T, C or G  
US-10-012-140-4  
Alignment Scores:  
Pred. No.: 1,14e-164 Length: 1820  
Score: 1348.00 Matches: 260  
Percent Similarity: 98.50% Conservative: 3  
Best Local Similarity: 97.38% Mismatches: 4  
Query Match: 97.33% Indels: 0  
Gaps: 0  
US-10-649-273-2\_copy\_148\_414 (1-267) x US-10-012-140-4 (1-1820)  
Qy 1 MetGlnAlaHisAlaLeuThrhIleargLeuThrhAsnlyValGluPheProPheLeuVal 20  
Db 587 ATGAGAGCTCATGACCTTACTATTAAGTTGACCAATAAAGTAGAATTTCTTTTACTT 646  
Qy 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValIleSerAspPheLeu 40  
Db 647 CTTTGTGATTTCTGGAGGTCACGTGCTGTGGCATTAGTTCAAGAGATTTCAGATTTTCTG 706  
Qy 41 LeuLeuGlyIySerLeuAspIleAlaProGlyAspMetLeuAspLyValAlaArgArg 60  
Db 707 CTTCTTGGAAGTCTTTGACATAGCACAGGTGACATGCTTTGACAGAGTGGCAAGAGA 766  
Qy 61 LeuSerLeuIlelyshisProgluCySerThrmeserlyglylysaalilegluhis 80  
Db 767 CTTCTTTTAAATAAATCAATCCAGAGTGTCCACCATGATGTGGGAAGCCATAGAACAT 826  
Qy 81 LeuAlaIyGlnGlysnatgphelaphaspilleyserproleuhsiasiallys 100  
Db 827 TTGGCCCAACAGAAATATGATTCATTTTACATCAAAACCTCCCTTCATCATGCTTAA 886  
Qy 101 AasnCyAspPheSerPheThrglyleuglnhisvalthrapslylleilemetlylys 120  
Db 887 AATGTGATTTTCTTTTACTGAGCTTCAACAGCTTACTGATTAATAAATATGAAAAACAG 946  
Qy 121 GlnyGlygluglylileglulysglylnlleuSerSerAlaIleaspilleala 140  
Db 947 AAACAAGAGAGATATGAGAGGGGCAATCTGTCTTCAGACAGACATTTGCTGCC 1006  
Qy 141 ThrValGlnhisThrmelAlaCyshisLeuVallysaThrhIsarGAlaIleleuphe 160  
Db 1007 ACAGTACAGACACACAATGGCATGTCATCTTGTGAAGAACACATCGGCTATTCTGTTT 1066  
Qy 161 CysIyGlnArgAspLeuProglInasnAsnAlaValleuValAlaSerGlyVal 180  
Db 1067 TGTAAAGCAGAGAGACTGTGATCTCAAAATATGACAGTACTGTGATCTGTGATGTC 1126  
Qy 181 AlaSerAspPheThrhIleargArgAlaLeuGlnlleuThrhAsnAlaThrhInCysThr 200  
Db 1127 GCAAGTAACTTCTATATCCGACAGCTCTGGAATTTTAAACAAGACACAGTCACT 1186  
Qy 201 LeuLeuCySerProProArgLeuCySerThrhAspAsnGlyllemetllealatrpsnGly 220  
Db 1187 TTGTTGTGTCCTCTCCACAGACTATGACATGATATGATGATGATGATGATGATGATGAT 1246  
Qy 221 IleGlnArgLeuArgAlaGlyleuglylleleuhsiaspillegluglylleargTyGlu 240  
Db 1247 ATTTGAAAGACTAGCTGCTGCTGGCATTTTACATGACATAGAGAGGATCCGCTATGAA 1306  
Qy 241 ProLyCySProlLeuGlyValAspIleSerLyGlnValGlylnAlaSerIlelyVal 260  
Db 1307 CCAAAATGTCCTTTGAGTAGACATATCAAAAGAGTGGAGAGCTTCATTAAGATA 1366  
Qy 261 ProGlnLeuIyMetGluIle 267  
Db 1367 CCACATTAATAATGGAGATA 1387

```
RESULT 15
US-10-094-749-400
; Sequence 400, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOMYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 400
; LENGTH: 2208
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-094-749-400

Alignment Scores:
Pred. No.:          9,06e-146      Length:          2208
Score:             1204.00         Matches:          239
Percent Similarity: 89.51%         Conservative:      0
Best Local Similarity: 89.51%       Mismatches:       4
Query Match:       86.93%          Indels:          24
DB:                17              Gaps:            1

US-10-649-273-2_COPY_148_414 (1-267) x US-10-094-749-400 (1-2208)

QY      1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB      785 ATGAGAGCTCATGCACTACTACTATTAGGTGACCAATTAAGTAGAATTTCCTTTTAACTT 844
QY      21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB      845 CTTTGAATTTCTGAGAGGTCACTGCTGTGTGGCATTAGTTCAGAGGTTTCAGATTTTCTG 904
QY      41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB      905 CTTCTTGAAAGGCTTTTGACATGACACAGGTGACATGCTTGACAGGTGGCAAGAGAA 964
QY      61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 80
DB      965 CTTCCTTAATAAACAATCCAGAGTGTCTCCACCATGAGTGTGGGAAAGCCATAGAACAT 1024
QY      81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
DB      1025 TTGGCCAAACAAGGAATAGATTTCATTTGACATCAAACTCCCTTCATCATCTATAA 1084
QY      101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
DB      1085 AATTGATATTTCCTTTACTGACCTTCACACGTTACTGATTAATAATGAAAAAG 1144
```

```
QY      121 GluLysGluGluGlyIleGlnLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
DB      1145 GAAAAAGAGGAAGGATTTGAGAAAGGGCAAAATCCCTGCTTCAGCAGACACTTCCTGCC 1204
QY      141 ThrValGlnHisThrMetAlaCysHisIleuValLysArgThrHisArgAlaIleLeuPhe 160
DB      1205 ACAGTACGACACACATGGCATGTCATCTTGAAAAAGAACACATCGGGCTAATCTGTTT 1264
QY      161 CysLysGlnArgAspLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
DB      1265 TGTAAAGCAGAGAGACTTGTACTTCAAAATTAATGACAGTACTGTCATCTGTGTGTC 1324
QY      181 AlaSerAspPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
DB      1325 GCAAGTAACTTCTGTATCCGACAGACTCTGGAAATTTTAAACAAACGCAACACAGTCACT 1384
QY      201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220
DB      1385 TTGTTGTCTCCTCCCTCCACAGACTATGCACTGATTAATGCAATTATGATTAAGCA 1435
QY      221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 240
DB      1435 ----- 1435
QY      241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260
DB      1436 ---TGATGTCTCTTGAGTACATATCAAAAGAGTTGGAGAACTTCCATAAAGTA 1492
QY      261 ProGlnLeuLysMetGluIle 267
DB      1493 CCACATTTAAAAATGAGATA 1513
```

Search completed: June 17, 2005, 08:03:19  
Job time : 1824.86 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: June 16, 2005, 15:19:45 ; Search time 2858.39 Seconds  
(without alignments)  
4051.513 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414

Perfect score: 1240  
Sequence: 1 LIALVQGVSDFLIKGSLDI.....DISKEVGEASIKVPQLNMEI 239

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 segs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seg length: 0  
Maximum DB seg length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool\_p/US10649273/runat\_15062005\_11416\_6030/app\_query.fasta\_1.1429  
-DB=GenEmbl -Qfmt=fastap -SUPFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-INITs=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10649273.@CEN\_1.1.7742.@runat\_15062005\_11416\_6030 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:.\*  
1: gb\_ba:.\*  
2: gb\_hcg:.\*  
3: gb\_in:.\*  
4: gb\_cm:.\*  
5: gb\_ov:.\*  
6: gb\_pat:.\*  
7: gb\_ph:.\*  
8: gb\_pl:.\*  
9: gb\_pr:.\*  
10: gb\_ro:.\*  
11: gb\_sts:.\*  
12: gb\_sy:.\*  
13: gb\_un:.\*  
14: gb\_vi:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Query Match | Length | DB ID           | Description        |
|------------|-------------|--------|-----------------|--------------------|
| 1          | 1240        | 100.0  | 1416 6 AR541929 | AR541929 Sequence  |
| 2          | 1240        | 100.0  | 1526 6 AR428809 | AR428809 Sequence  |
| 3          | 1240        | 100.0  | 1908 9 BC011904 | BC011904 Homo sapi |
| 4          | 1240        | 100.0  | 2197 6 AR428803 | AR428803 Sequence  |

|    |        |      |        |    |           |                    |
|----|--------|------|--------|----|-----------|--------------------|
| 5  | 1213.5 | 97.9 | 1387   | 6  | AR428808  | AR428808 Sequence  |
| 6  | 1213.5 | 97.9 | 1387   | 9  | HSN295148 | AJ295148 Homo sapi |
| 7  | 1203   | 97.0 | 1245   | 6  | AX664697  | AX664697 Sequence  |
| 8  | 1203   | 97.0 | 1820   | 6  | AX664695  | AX664695 Sequence  |
| 9  | 1086   | 87.6 | 1844   | 10 | BC058172  | BC058172 Mus muscu |
| 10 | 1079   | 87.0 | 1017   | 10 | BC038910  | BC038910 Mus muscu |
| 11 | 1059   | 85.4 | 2208   | 6  | AX713716  | AX713716 Sequence  |
| 12 | 1059   | 85.4 | 2208   | 9  | AK055441  | AK055441 Homo sapi |
| 13 | 983    | 79.3 | 1546   | 10 | BC078974  | BC078974 Rattus no |
| 14 | 944.5  | 76.2 | 1109   | 10 | BC051211  | BC051211 Mus muscu |
| 15 | 809    | 65.2 | 860    | 5  | BX930694  | BX930694 Gallus ga |
| 16 | 806    | 65.0 | 1522   | 5  | BX934991  | BX934991 Gallus ga |
| 17 | 803    | 64.8 | 1558   | 5  | BX930963  | BX930963 Gallus ga |
| 18 | 725    | 58.5 | 14364  | 6  | AR428807  | AR428807 Sequence  |
| 19 | 725    | 58.5 | 84115  | 9  | AC013468  | AC013468 Homo sapi |
| 20 | 666    | 53.7 | 249601 | 2  | AC114153  | AC114153 Rattus no |
| 21 | 666    | 53.7 | 308652 | 2  | AC121478  | AC121478 Rattus no |
| 22 | 646.5  | 52.1 | 256751 | 10 | AC122925  | AC122925 Mus muscu |
| 23 | 610    | 49.2 | 875    | 6  | CQ721898  | CQ721898 Sequence  |
| 24 | 384.5  | 31.0 | 1474   | 3  | AK113378  | AK113378 Clona int |
| 25 | 369.5  | 29.8 | 121251 | 5  | AL591593  | AL591593 Zebrafish |
| 26 | 356    | 28.7 | 20021  | 10 | AF367967  | AF367967 Mus muscu |
| 27 | 356    | 28.7 | 179252 | 10 | AF131205  | AF131205 Mus muscu |
| 28 | 345    | 27.8 | 117322 | 5  | AL672217  | AL672217 Zebrafish |
| 29 | 332.5  | 26.8 | 1576   | 3  | AY051882  | AY051882 Drosophi  |
| 30 | 332.5  | 26.8 | 1601   | 6  | CQ606432  | CQ606432 Sequence  |
| 31 | 333.5  | 26.8 | 3656   | 6  | CQ606431  | CQ606431 Sequence  |
| 32 | 332.5  | 26.8 | 14679  | 2  | AC018262  | AC018262 Drosophi  |
| 33 | 332.5  | 26.8 | 180263 | 3  | AC010671  | AC010671 Drosophi  |
| 34 | 332.5  | 26.8 | 207432 | 3  | AE003513  | AE003513 Drosophi  |
| 35 | 324    | 26.1 | 1443   | 8  | AY024338  | AY024338 Arabidops |
| 36 | 324    | 26.1 | 1474   | 8  | AY117283  | AY117283 Arabidops |
| 37 | 324    | 26.1 | 1567   | 8  | AY063864  | AY063864 Arabidops |
| 38 | 322    | 26.0 | 1557   | 8  | AY084577  | AY084577 Arabidops |
| 39 | 311.5  | 25.1 | 1672   | 8  | AK070912  | AK070912 Oryza sat |
| 40 | 305    | 24.6 | 110000 | 2  | AP065011  | AP065011 Continuat |
| 41 | 262.5  | 21.2 | 333800 | 1  | SM591792  | SM591792 Sinorhizo |
| 42 | 260    | 21.0 | 1718   | 8  | AK099665  | AK099665 Oryza sat |
| 43 | 257    | 20.7 | 349354 | 1  | BX640416  | BX640416 Bordetell |
| 44 | 254    | 20.5 | 260271 | 1  | AE017258  | AE017258 Molbschia |
| 45 | 253    | 20.4 | 349116 | 1  | AP003003  | AP003003 Mesorhizo |

#### ALIGNMENTS

RESULT 1  
AR541929 LOCUS AR541929 1416 bp DNA linear PAT 08-OCT-2004  
DEFINITION Sequence 177 from patent US 6743619.  
ACCESSION AR541929  
VERSION AR541929.1 GI:53934009  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1416)  
AUTHORS Tang,Y.T., Zhou,P., Goodrich,R., Liu,C., Asundi,V., Ren,P.,  
Zhang,J., Zhao,Q.A., Yang,Y., Xue,A.D., Wehman,T., Wang,J.-R.,  
Wang,D. and Drmanac,R.T.  
TITLE Nucleic acids and polypeptides  
JOURNAL Patent: US 6743619-A 177 01-JUN-2004;  
FEATURES  
source location/Qualifiers  
1..1416  
/organism="unknown"  
/mol\_type="genomic DNA"

#### ORIGIN

Alignment Scores:  
Pred. No.: 2,246-113  
Score: 1240.00 Length: 1416  
Percent Similarity: 100.00% Matches: 239  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0



DB: 6 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x AR541929 (1-1416)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
 |||||  
 DB 586 CTGTGGCACTGATGTTGAAGAGATTTCAGATTTCCTGTTTGGAAAGCTTTGGACATA 645

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
 |||||  
 DB 646 GCACCAAGTGACATGCTTGAACAGGTGGCAAGAAAGCTTTCTTAATAAATCCAGAG 705

QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
 |||||  
 DB 706 TGCTCCACCATGATGTTGGTGGAAAGCCATAGAACATTGGCCAAACAGAAATAGATT 765

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80  
 |||||  
 DB 766 CATTTTGACATCAAACTCCCTTGCAATCATGCTAAATAATGTGATTTTCTTTTACTGA 825

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIleGluLys 100  
 |||||  
 DB 826 CTTCACACGTTACTGATTAATAATATATGAAAAAGAAAAAGGAAAGATTTGAGAG 885

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120  
 |||||  
 DB 886 GGGCAATATCTGCTTCCAGACAGCATTTGCTGCCACATGACACACATGGCATGT 945

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
 |||||  
 DB 946 CATCTGTGAAAAAGAACACATCGGGCTATTCTGTTTGTAAAGACAGAGACTTGTACT 1005

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaAspSerAspPheTrpIleArgArg 160  
 |||||  
 DB 1006 CAAATAATGCGAGTACTGTGTCATCTGTGTGTCAGAGTAATCTTAATCCGAGA 1065

QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
 |||||  
 DB 1066 GCTGTGAAATTTTAACAAACGACACAGTGCATTTGTGTCTCTCTCCAGACTA 1125

QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
 |||||  
 DB 1126 TGACACTGATTAATGCGATTAATGATTCATGGAATGATTAAGACATGCTGCTTG 1185

QY 201 GlyIleLeuHisAspIleGluGlyIleArgTrpGluProLysCysProLeuGlyValAsp 220  
 |||||  
 DB 1186 GGCATTATTACATGACATAGAGGCAATCCGCTATGAACCAAAATGTCTCTTGGAGTAGAC 1245

QY 221 HisSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239  
 |||||  
 DB 1246 ATATCAAAAGAGTTGAGAGAGCTTCATTAAGTACCAATTAATAATGAGATA 1302

RESULT 2  
 AR428809 1526 bp DNA linear PAT 18-DEC-2003  
 LOCUS Sequence 23 from patent US 6642041.  
 DEFINITION AR428809  
 ACCESSION AR428809  
 VERSION AR428809.1 GI:40188595  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 1526)  
 AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.  
 TITLE Polynucleotides encoding a novel metalloprotease, MP-1  
 JOURNAL Patent: US 6642041-A 23 04-NOV-2003;  
 FEATURES  
 source 1..1526  
 /organism="Unknown"  
 /mol\_type="genomic DNA"

Pred. No.: 2,456-113 Length: 1526  
 Score: 1240.00 Matches: 239  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x AR428809 (1-1526)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
 |||||  
 DB 85 CTGTGGCACTGATGTTGAAGAGATTTCAGATTTCCTGTTTGGAAAGCTTTGGACATA 144

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
 |||||  
 DB 145 GCACCAAGTGACATGCTTGAACAGGTGGCAAGAAAGCTTTCTTAATAAATCCAGAG 204

QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
 |||||  
 DB 205 TGCTCCACCATGATGTTGGTGGAAAGCCATAGAACATTGGCCAAACAGAAATAGATT 264

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80  
 |||||  
 DB 265 CATTTTGACATCAAACTCCCTTGCAATCATGCTAAATAATGTGATTTTCTTTTACTGA 324

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIleGluLys 100  
 |||||  
 DB 325 CTTCACACGTTACTGATTAATAATATATGAAAAAGAAAAAGGAAAGATTTGAGAG 384

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120  
 |||||  
 DB 385 GGGCAATATCTGCTTCCAGACAGCATTTGCTGCCACATGACACACATGGCATGT 444

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
 |||||  
 DB 445 CATCTGTGAAAAAGAACACATCGGGCTATTCTGTTTGTAAAGACAGAGACTTGTACT 504

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaAspSerAspPheTrpIleArgArg 160  
 |||||  
 DB 505 CAAATAATGCGAGTACTGTGTCATCTGTGTGTCAGAGTAATCTTAATCCGAGA 564

QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
 |||||  
 DB 565 GCTGTGAAATTTTAACAAACGACACAGTGCATTTGTGTCTCTCTCCAGACTA 624

QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
 |||||  
 DB 625 TGACACTGATTAATGCGATTAATGATTCATGGAATGATTAAGACATGCTGCTTG 684

QY 201 GlyIleLeuHisAspIleGluGlyIleArgTrpGluProLysCysProLeuGlyValAsp 220  
 |||||  
 DB 685 GGCATTATTACATGACATAGAGGCAATCCGCTATGAACCAAAATGTCTCTTGGAGTAGAC 744

QY 221 HisSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239  
 |||||  
 DB 745 ATATCAAAAGAGTTGAGAGAGCTTCATTAAGTACCAATTAATAATGAGATA 801

RESULT 3  
 BC011904 1908 bp mRNA linear PRI 23-DEC-2003  
 LOCUS Homo sapiens O-6-ialoglycoprotein endopeptidase-like 1, mRNA (cDNA  
 clone MGC:20293 IMAGE:4121450), complete cds.  
 DEFINITION BC011904  
 ACCESSION BC011904  
 VERSION BC011904.2 GI:40225818  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 1908)  
 Straubeberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,  
 Klausner,R.D., Collins,F.S., Wagner,L., Schmen,C.M., Schuler,G.D.,  
 Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,

Alignment Scores:

| JOURNAL   | TITLE   | REMARK   |
|-----------|---|--|
| PUBMED    | 12477932  | NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a> |
| REFERENCE | 2 (bases 1 to 1908)   | On Dec 19, 2003 this sequence version replaced gi:15080281.                      |
| REFERENCE | Strausberg, R.  | Contact: MGC help desk   |
| AUTHORS   | Direct Submission   | Email: <a href="mailto:gcgaps-remail.nih.gov">gcgaps-remail.nih.gov</a>          |
| TITLE     | Submitted (30-JUL-2001) National Institutes of Health, Mammalian                                  | Tissue Procurement: ATCC   |
| JOURNAL   | Gene Collection (MGC), Cancer Genomics Office, National Cancer                                    | CNA Library Preparation: Rubin Laboratory  |
|           | Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,                                  | cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)                         |
|           | USA   | DNA Sequencing by: National Institutes of Health Intramural                      |
|           |   | Sequencing Center (NISC),  |
|           |   | Gaithersburg, Maryland.  |
|           |   | Web site: <a href="http://www.nisc.nih.gov/">http://www.nisc.nih.gov/</a>        |
|           |   | Contact: <a href="mailto:nisc@ncihgri.nih.gov">nisc@ncihgri.nih.gov</a>          |
|           |   | Ahter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,                   |
|           |   | Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,            |
|           |   | Dietch, N.L., Granite, S., Gan, X., Gupta, J., Haghighi, P.,                     |
|           |   | Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R.,           |
|           |   | Maduro, Q.L., Masiello, C., Makkeri, B., Mastrian, S.D., McCloskey, J.C.,        |
|           |   | McDowell, J., Pearson, R., Statistik, S., Thomas, P.J., Touchman, J.W.,          |
|           |   | Taurigon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,             |
|           |   | Young, A., Zhang, L. H. and Green, E.D.  |
| FEATURES  | Clone distribution: MGC clone distribution information can be found                               |  |
| source    | through the I.M.A.G.E. Consortium/LNL at: <a href="http://image.lnl.gov">http://image.lnl.gov</a> |  |
|           | Series: IRAL Plate: 28 Row: 1 Column: 22.   |  |
|           | Location/Qualifiers   |  |
|           | 1. 1908   |  |
|           | /organism="Homo sapiens"  |  |
|           | /mol_type="mRNA"  |  |
|           | /db_xref="taxon:9606"   |  |
|           | /clone="MGC:20293 IMAGE:4121450"  |  |
|           | /tissue_type="Muscle, rhabdomyosarcoma"   |  |
|           | /clone_lib="NIH MGC 17"   |  |
|           | /lab_host="DH10B-R"   |  |
|           | /note="Vector: pOTB7"   |  |
| gene      | 1. 1908   |  |
|           | /gene="OSGRL1"  |  |
|           | /db_xref="locusID:64172"  |  |
|           | 104 1348  |  |
|           | /gene="OSGRL1"  |  |
|           | /codon_start=1  |  |
|           | /product="OSGRL1 protein"   |  |
|           | /protein_id="AAH11904.1"  |  |
|           | /db_xref="GI:15080281"  |  |
|           | /db_xref="locusID:64172"  |  |
|           | /translation="MLITRTAGCFFPPSRKRYEPLRSRFRHGTLPFKIVGIGIT  |  |
|           | GVSDSDSAIVATIKPGLALSHGSEFHLTKGIVPPAQQVLRNIRIVQELASAS  |  |
|           | GCSTSDSAIVATIKPGLALSHGSEFHLTKGIVPPAQQVLRNIRIVQELASAS  |  |

misc\_feature  
215..1111  
/gene="OSGEPL1"  
/note="Peptidase M82; Region: Glycoprotease family"  
/db\_xref="CD:pfam00814"

ORIGIN

Alignment Scores:

| Pred. No.:             | 3.19e-113 | Length:       | 1908 |
|------------------------|-----------|---------------|------|
| Score:                 | 1240.00   | Matches:      | 239  |
| Percent Similarity:    | 100.00%   | Conservative: | 0    |
| Best Local Similarity: | 100.00%   | Mismatches:   | 0    |
| Query Match:           | 100.00%   | Indels:       | 0    |
| DB:                    | 9         | Gaps:         | 0    |

US-10-649-273-2\_COPY\_176\_414 (1-239) x BC011904 (1-1908)

|    |      |   |      |
|----|------|---|------|
| QY | 1    | LeuLeuAlaLeuValGlnGlyValIserAspPheLeuLeuGlyValSerLeuAspIle      | 20   |
| Db | 629  | CTGTGTGGCATTAAGTTCACAGAGATTTCAGATTTTCCTCTTGGAAAGCTTTGGACATA     | 688  |
| QY | 21   | AlAProGlyAspMetLeuAspIleValAlaArgArgLeuSerLeuIleIysHisProGlu    | 40   |
| Db | 689  | GCACAGGTGACATGCTTGACAGGTGGCAAGAGACTTTCTTTATATAAACATCCAGAG       | 748  |
| QY | 41   | CysSerThrMetSerGlyGlyValenAlaIleGlnHisLeuAlaIysGlnGlyAsnArgPhe  | 60   |
| Db | 749  | TGCTCCACCATGAGGTGGTGGAAAGCCATTAAACATTTGGCCAAACAGAAATGATTT       | 808  |
| QY | 61   | HisPheAspIleIysProProLeuHisHisAlaIysAsnCyAspPheSerPheThrGly     | 80   |
| Db | 809  | CATTGTGACATCAAACTCCCTGCATCATCTAAAAATGTGATTTTCTTTTAACTGGA        | 868  |
| QY | 81   | LeuGlnHisValThrAspIleIleIleMetIysGlyGlyValGlnGlyIleGlnIys       | 100  |
| Db | 869  | CTTCACACGTTACGATTAATATATATATGAAAAAGAAAAAGAGATTTGAGAGAG          | 928  |
| QY | 101  | GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys | 120  |
| Db | 929  | GGGCAATTCCTGTCTTCAGCAGCAGACATTCCTGCCACAGTACAGACACATGGCATGT      | 988  |
| QY | 121  | HisLeuValIysArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuLeuPro    | 140  |
| Db | 989  | CATCTGTGAAAAAGAACACATCGGGCTATTCTGTTTGTAAAGCAGAGACTTGTTACTCT     | 1048 |
| QY | 141  | GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaIleSerAsnPheIleIleArgArg | 160  |
| Db | 1049 | CAAAATATATGACGACTGCTGTCGATCTGCTGTGTCGCAAGATTAATCTTATATCCGACGA   | 1108 |
| QY | 161  | AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrIleLeuLeuCysProProProArgLeu | 180  |
| Db | 1109 | GCTCTGGAATTTTAAACAAACGCAACACAGTGCATTTGTGTGTGTCTCTCTCCAGACTA     | 1168 |
| QY | 181  | CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeu | 200  |
| Db | 1169 | TGCATCTGATTAATGCGATTATGATTCGATCGAATGTATTTGAAAGACTACGTCGTGGCTTG  | 1228 |
| QY | 201  | GlyIleLeuHisAspIleGlnGlyIleArgGlyArgGluProIysCysProLeuGlyValAsp | 220  |
| Db | 1229 | GGCATTTTATACATGACATAGAGGCGATCCGCTATGACCAAAATATGCTCTCTGAGTTAGC   | 1288 |
| QY | 221  | IleSerIysGlnValGlyGlnAlaSerIleIysValProGlnIleIysMetGlnIle       | 239  |
| Db | 1289 | ATATCAAAAGAGATTGGAGAGCTTCCATPAAAAGTACCACAATTAAAAATGAGAGATA      | 1345 |

ACCESSION AR428803  
VERSION AR428803.1 GI:40188589  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
1 (bases 1 to 2197)  
AUTHORS Chen, J., Feder, J.N., Nelson, T.C., Krystek, S.R. and Duclos, F.  
TITLE Polynucleotides encoding a novel metalloprotease, MP-1  
JOURNAL Patent: US 6642041-A 1 04-NOV-2003;  
FEATURES Location/Qualifiers  
source 1..2197  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Alignment Scores:  
Pred. No.: 3,77e-113 Length: 2197  
Score: 1240.00 Matches: 239  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x AR428803 (1-2197)

1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
756 CTGTTGGCATTAGTTCAGAGAGTTTCAGATTTCCTTCTTGAAAGCTCTTGAGACATA 815  
21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
816 GCACCAAGTGACATGCTTGCAAGTGCGCAAGAAGCTTTCTTAATAAATCCAGAG 875  
41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaLysGInGlyAsnArgPhe 60  
876 TCGTCCACCATGATGTTGGTGGAAAGCCATAGCAATTGGCCAAACAGGAATAGATT 935  
61 HisPheAspIleLysProProLeuHisAlaLysAsnGlyAspPheSerPheThrGly 80  
936 CATTTCGACATCAACCTCCCTTCGATCATGCTTAATAATGTGATTTTCTTTTACTGGA 995  
81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIleGluLys 100  
996 CTTCAACAGCTTACTGATTAATAATATATGAAAAAGAAAAAGAGATATTGAGAG 1055  
101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120  
1056 GGGCAATATCCTGCTTCAGCAGCAGACATGCTGCCACAGTACAGCACCAATGCGATGT 1115  
121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
1116 CATCTTGTAAGAAAGAACATCGGGCTATCTGTTTGTAAAGCAGAGACTGTACT 1175  
141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaLaserAsnPheTyrlleArgArg 160  
1176 CAATAAATAGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1235  
161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
1236 GCTCTGAATTTTAAACAAGCAGACACATGCACTTTGTTGTCTCTCCCAAGCTA 1295  
181 CysThrAspAsnGlyIleMetIleAlaTPAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
1296 TGCACATGATATGCAATATGATGATGATGATGATGATGATGATGATGATGATGATG 1355  
201 GlyIleLeuHisAspIleGluGlyIleArgTyrlleProLysCysProLeuGlyValAsp 220  
1356 GGCATTTTACATACATAGAGGCAATCGCTATGAAACCAAAATGCTCTCTTGAGTAC 1415  
221 IleSerLysGluValGlyAlaSerIleLysValProGlnLeuLysMetGluIle 239  
1416 AATATCAAAAGAACTTGGAGAGCTTCATAAAGTACACATTTAAATAAGAGATA 1472

RESULT 5  
AR428808  
LOCUS AR428808 1387 bp DNA linear PAT 18-DEC-2003  
DEFINITION Sequence 21 from patent US 6642041.  
ACCESSION AR428808  
VERSION AR428808.1 GI:40188594  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
1 (bases 1 to 1387)  
AUTHORS Chen, J., Feder, J.N., Nelson, T.C., Krystek, S.R. and Duclos, F.  
TITLE Polynucleotides encoding a novel metalloprotease, MP-1  
JOURNAL Patent: US 6642041-A 21 04-NOV-2003;  
FEATURES Location/Qualifiers  
source 1..1387  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Alignment Scores:  
Pred. No.: 9.37e-111 Length: 1387  
Score: 1213.50 Matches: 238  
Percent Similarity: 90.15% Conservative: 0  
Best Local Similarity: 90.15% Mismatches: 1  
Query Match: 97.86% Indels: 25  
Gaps: 1

US-10-649-273-2\_COPY\_176\_414 (1-239) x AR428808 (1-1387)

1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
549 CTGTTGGCATTAGTTCAGAGAGTTTCAGATTTCCTTCTTGAAAGCTCTTGAGACATA 608  
21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
609 GCACCAAGTGACATGCTTGCAAGTGCGCAAGAAGCTTTCTTAATAAATCCAGAG 668  
41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaLysGInGlyAsnArgPhe 60  
669 TCGTCCACCATGATGTTGGTGGAAAGCCATAGCAATTGGCCAAACAGGAATAGATT 728  
61 HisPheAspIleLysProProLeuHisAlaLysAsnGlyAspPheSerPheThrGly 80  
729 CATTTCGACATCAACCTCCCTTCGATCATGCTTAATAATGTGATTTTCTTTTACTGGA 788  
81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIle----- 98  
789 CTTCAACAGCTTACTGATTAATAATATATGAAAAAGAAAAAGAGATATTGCTA 848  
98 ----- 98  
849 ATTAGTAAAGTTGAACAGATTAATATTCCTGATTTGCTTAATAATAGCTGCTATTTC 908  
99 -----GluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGln 115  
909 TGCAGTATGAGAGGGGCAATCTGCTTCAGAGCAGACATGCTGCCACAGTACAG 968  
116 HisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGln 135  
969 CACCAATGCGATGCTCTGTTGAAAAAGAACATCGGGCTATCTGTTTGTATAGCAG 1028  
136 ArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAlaLaserAsn 155  
1029 AGAGACTTTGTAACCTCAAAATAATAGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTG 1088  
156 PheTyrlleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCys 175  
1089 TTCTATATCGCAGAGCTCTGGAATTTTAAACAAGCAGACAGTGCATTTGTTGT 1148  
176 ProProArgLeuGluCysThrAspAsnGlyIleMetIleAlaTPAsnGlyIleGluArg 195

|  |   |  |                             |
|--|---|--|-----------------------------|
| Db                                     | 1149  | CCTCTCCCAAGACATATGACATGATTAATGGCATTTATGATTCATGGAATGATTTAAAGA | 1208                        |
| Qy                                     | 196   | LeuA   | 196                         |
| Db                                     | 1209  | CTAGTGGTGGCTTGGGCGATTTTACATGACATGAGGATCCGCTATGAACAAAATGT     | 1268                        |
| Qy                                     | 216   | ProLeuG  | 216                         |
| Db                                     | 1269  | CCTTTGGAGTGAACATATCAAAAGAGTTGGAGAGCTTCATTAAGTACCAATTA        | 1328                        |
| Qy                                     | 236   | lysMetG  | 236                         |
| Db                                     | 1329  | AAATGAGATTA  | 1340                        |
| RESULT 6                               |   |  |                             |
| LOCUS                                  | HS295148  | 1387 bp  | mRNA linear PRI 30-OCT-2000 |
| DEFINITION                             | Homo sapiens mRNA for putative sialoglycoprotease type 2.   |  |                             |
| ACCESSION                              | AJ295148  |  |                             |
| VERSION                                | AJ295148.1  | GI:11071726  |                             |
| KEYWORDS                               | metallopeptidase; sialoglycoprotease.   |  |                             |
| SOURCE                                 | Homo sapiens (human)  |  |                             |
| ORGANISM                               | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.   |  |                             |
| REFERENCE                              | 1   | Chen, J.M., Fortunato, M. and Barrett, A.J.                  |                             |
| AUTHORS                                | Cloning and sequencing of a second human putative sialoglycoprotease homologue  |  |                             |
| TITLE                                  | 2 (bases 1 to 1387)   |  |                             |
| JOURNAL                                | Unpublished   |  |                             |
| AUTHORS                                | Chen, J.M.  |  |                             |
| REFERENCE                              | Direct Submission   |  |                             |
| AUTHORS                                | Submitted (27-OCT-2000) Chen J.M., MRC Molecular Enzymology Laboratory, The Babraham Institute, Babraham, Cambridge, CB2 4AT, UNITED KINGDOM  |  |                             |
| JOURNAL                                | UNITED KINGDOM  |  |                             |
| FEATURES                               |   |  |                             |
| SOURCE                                 | location/Qualifiers   |  |                             |
|  | 1..1387   |  |                             |
|  | /organism="Homo sapiens"  |  |                             |
|  | /mol_type="mRNA"  |  |                             |
|  | /db_xref="taxon:9606"   |  |                             |
|  | /tissue_type="uterus"   |  |                             |
|  | 24..1343  |  |                             |
|  | /function="metallopeptidase of family M22"  |  |                             |
|  | /codon_start=1  |  |                             |
|  | /product="putative sialoglycoprotease type 2"   |  |                             |
|  | /protein_id="CAC14666.1"  |  |                             |
|  | /db_xref="GI:11071727"  |  |                             |
|  | /db_xref="GOA:Q9H4B0"   |  |                             |
|  | /db_xref="UniProt/TTrEMBL:Q9H4B0"   |  |                             |
|  | /translation="MLITTKAGVFPKSKRYVEFLRSFNPHEPTLFLAKIVIGIETSCDDTAADVDEGTGVLGEALHSQTEVHLKTCGVLPAPAOOLHRENIQRIVOEALASAGSDDSDSAIAITTKIPKGLSLGVSTSLQVLQKPFPIHMEAHALTRITNNVVEPPVIVILSGCHCLALVGVSDPFLITGSDIDAPDMDVARRSLIKHPKESCTMSGKALBHLAKQGRFRFPDIKPLTHAKNCDFSTGYQHTYDKIINKEGEBSIFPLSKVQDINIPGLCLAKIAHFCKRYEKQILSSADIDATVQHTMACHLVKTRERAILPCKQRDLIPNNNAVIVASGVASNFYIRRALEIITNATQCTLCPPPLCTDNGIMTAMNGIIERLRGIGILIHIDIEGIRYBPCKPLGVDISKEVGEASIKVQLKMEI" |  |                             |
| ORIGIN                                 |   |  |                             |
| Alignment Scores:                      |   |  |                             |
| Pred. No.:                             | 9.37e-111   | Length:  | 1387                        |
| Score:                                 | 1213.50   | Matches:   | 238                         |
| Percent Similarity:                    | 90.15%  | Conservative:  | 0                           |
| Best Local Similarity:                 | 90.15%  | Mismatches:  | 1                           |
| Query Match:                           | 97.86%  | Indels:  | 25                          |
| Db:                                    | 9   | Gaps:  | 1                           |
| US-10-649-273-2_COPY_176_414 (1-239) x | HS295148 (1-1387)   |  |                             |
| Qy                                     | 1   | LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIle | 20                          |
| Db                                     | 549   | CTGTGGCAATTAAGTTCAGAGATTCACAGATTTTCGCTCTTGGAAGCTTTTGACATTA   | 608                         |

|            |   |  |        |
|------------|---|--|--------|
| Oy         | 21  | AlaProGlyAspMetCLeuApIysValAlaArgArgLeuSerLeuIleYshIProGlu     | 40     |
| Db         | 609   | GCACCAAGTGCATGCTTGCACAGGTGGCAAGAAGACTTTCTTTAAATCAATCCAGAG      | 668    |
| Oy         | 41  | CysSerThrMetSerGlyIleYysAlaIleGluHisLeuAlaIysGlnIleYasnArgPhe  | 60     |
| Db         | 669   | TGCTCCACCATGATGGTGGGAAAGCCATGAGCATTTGGCCAAAGAAATATGATTTT       | 728    |
| Oy         | 61  | HisPheAspIleYysProProLeuHisHisAlaIaYasnCybAspPheSerPheThrGly   | 80     |
| Db         | 729   | CATTTGACATCAAAACCTCCCTGCATCAGCTCAAAAATGTGTATTTTCTTTTACTGGA     | 788    |
| Oy         | 81  | LeuGlnHisValThrAspYsIleIleMetIleYysGlnYysGlnGluGlnIle-----     | 98     |
| Db         | 789   | CTTCAACACGTTACTGATATAAATATATATGAAAAAGAAAAAGAGAAAGATATATTTCTA   | 848    |
| Oy         | 98  | -----  | 98     |
| Db         | 849   | ATTATGTAAGTTGAACGATTAATATTCCTGGATTTGCTCCCTAAAAATAGCTGCATTTTC   | 908    |
| Oy         | 99  | -----GluYysGlyGlnIleLeuSerSerAlaIaAspIleAlaIaThrValGln         | 115    |
| Db         | 909   | TGCAGATATAGAGAGGGGCAAAATCTGTCTTCACAGACAGCATTTGCTGCCACAGTACAG   | 968    |
| Oy         | 116   | HisThrMetAlaCybHisLeuValIysArgThrHisArgAlaIleLeuPheCysYysGln   | 135    |
| Db         | 969   | CACACAATGGCATGTCAATCTTGTAAGAAAGAACATCGGGCTATCTGTGTTTGTAAACAG   | 1028   |
| Oy         | 136   | ArgAspLeuLeuProGlnAsnAsnAlaValIleuValIaSerGlyIleValIaSerAsn    | 155    |
| Db         | 1029  | AGAACACTGTTAATCCTCAAAATAATAGCAGTACTGTGTGATCTGGGGTGTGCAAGTAAAC  | 1088   |
| Oy         | 156   | PheTyrIleArgArgAlaLeuGlnIleLeuHisAsnAlaThrGlnCysThrLeuLeuCys   | 175    |
| Db         | 1089  | TTCTATATCCGACAGAGCTCTGGAATTTTAAACAACCAACACAGCTGACTTTGTGTGT     | 1148   |
| Oy         | 176   | ProProProArgLeuCybThrAspAsnGlyIleMetIleAlaTProAsnGlyIleGluArg  | 195    |
| Db         | 1149  | CCTCCTCCAGACTATGACATGATATGGCATTTATGTGATTCGATGAATGTATTTGAAGA    | 1208   |
| Oy         | 196   | LeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGluProIysCys   | 215    |
| Db         | 1209  | CTACAGTGTGGCTTGGGCATTTTTCATGACATTAAGAGCATCCGCTATGAACCAAAATGT   | 1268   |
| Oy         | 216   | ProLeuGlyValAspIleSerIysGluValAlaGlyGlnIaSerIleIysValProGluLeu | 235    |
| Db         | 1269  | CCTCTGGAGTAGACATATCAAAAGAAAGTGGAGAGCTTCATTAAGTACCAACAAATTA     | 1328   |
| Oy         | 236   | LysMetGluIle   | 239    |
| Db         | 1329  | AAATATGAGATTA  | 1340   |
| RESULT 7   |   |  |        |
| LOCUS      | AX664697  | 1245 bp  | DNA    |
| DEFINITION | Sequence 6 from Patent WO020749960.                                 |  | linear |
| ACCESSION  | AX664697  |  |        |
| VERSION    | AX664697.1  | GI:29164457  |        |
| KEYWORDS   |   |  |        |
| SOURCE     |   |  |        |
| ORGANISM   | Homo sapiens (human)  |  |        |
|            | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;   |  |        |
|            | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.           |  |        |
| REFERENCES | 1   |  |        |
| AUTHORS    | leibY, K.R., Kapeller-libermann,R. and Glucksmann,M.                |  |        |
| TITLE      | 38650, 28672, 5495, 65507, 81588 and 14354 methods and compositions |  |        |
| JOURNAL    | of human proteins and uses thereof                                  |  |        |
| FEATURES   | Patent: WO 02074960-A 6 26-Sep-2002;                                |  |        |
|            | Millennium Pharmaceuticals, Inc. (US)                               |  |        |
|            | Location/Qualifiers   |  |        |
|            | 1..1245   |  |        |
|            | /organism="Homo sapiens"  |  |        |

ORIGIN /mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

## Alignment Scores:

Pred. No.: 9,1e-110 Length: 1245  
Score: 1203.00 Matches: 232  
Percent Similarity: 98.33% Conservative: 3  
Best Local Similarity: 97.07% Mismatches: 4  
Query Match: 97.02% Indels: 0  
Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x AX664697 (1-1245)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
Db CTTGTGGCACTTACTTCAAGAGAGTTTCAGATTCTGCTTCTTGGAAAGCTTTGGACATA 585  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
Db GCAACGACGAGACATGCTTGCACAGGTGGCAAGAACATCTTTAATTAATAACATCCAGAG 645  
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60  
Db TGCTCCACATGAGTGTGGTGGAAAGCCATAGAACATTGGCCAAACAGAAATAGATT 705  
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
Db CATTGTGACATCAAACTCCCTTGCACTGTCATGCTTAATAATGTGATTTCTTTTACGGA 765  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysGlyLysGlnGlyLysGlnGlyLys 100  
Db CTTCAACAGCTTACTGATTAATAATAATGAAACAGGAAACAGAGGAGATTGAGAG 825  
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisSerMetAlaCys 120  
Db GGGCAAAATCTGCTTCCAGACAGACATTTGCTGCCACAGTACAGCACAAATGGCATGT 885  
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
Db CATCTTGTGAAAAGAACATCGGGCTATTCTGTTTGTGAGCAGAGACCTTGTACT 945  
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaLysSerAsnPheTyrlleArgArg 160  
Db CAAATAATGACAGTACTGCTGTCATCGTGCTGTCAGTACCTTATATCCGAGA 1005  
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180  
Db GCTCTCGAAATTTTAAACAAACGACACAGTGCATTTGTGTGCTCTCCACAGACTA 1065  
QY 181 CysThrAspAsnGlyIleMetIleAlaIlePheGlnGlyIleGlnArgLeuArgAlaGlyLeu 200  
Db TGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1125  
QY 201 GlyIleLeuHisAspIleGlnGlyLysArgTyrlleArgTyrlleArgTyrlleArgTyrlle 220  
Db GGCATTTTACATACATACATACATACATCCGCTATGAAACAAATCTCTTTCGAGTGAC 1185  
QY 221 HisSerLysGlyValGlyGlnAlaSerIleLysValProGlnLeuLysMetGlnIle 239  
Db 1186 AATATCAAAAGAGTTGAGAGAGCTTCATTAAGTACCAATTAATAATGAGAGATA 1242

RESULT 8  
AX664695  
LOCUS AX664695 1820 bp DNA linear PAT 22-MAR-2003  
DEFINITION Sequence 4 from Patent WO2074960.  
ACCESSION AX664695  
VERSION AX664695.1 GI:29164455  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1  
Leiby, K.R., Kapeller-Libermann, R. and Glucksmann, M.  
AUTHORS 38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions  
TITLE of human proteins and uses thereof  
JOURNAL Patent: WO 02074960-A 4 26-SEP-2002;  
Milleium Pharmaceuticals, Inc. (US)  
FEATURES Location/Qualifiers

## source

1..1820  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
146..1390  
/note="unnamed protein product"

## CDS

/codon\_start=1  
/protein\_id="CAD80044.1"  
/translation="MLITRTAGVFPFKRSKRYEPLRFSEFHPGLTFLHKIVGLET  
SCDDTAAAVDETENVGEALHSQTEVHLKTKGIVPPAQQLEHNTORIYQNALAS  
GVSPSDLSAIAITTKPGALSLVGLSPLOLVGLKPEPIHHEMHALITRLTK  
VEPFLVLISGHCILALVOGVSDFLKSLDIARDMDLAKNARLSLKHPECT  
MSGKALIEHLAKQGNRFPDIKPLHAKNCDSPFTGLQHTDKNENRKBEGIEKG  
OILSSADIAIVQHTMAHLVKRTHRAILPCXORDLLPQNNAVLVASGVAANFYIR  
RALEILTMNQCTLICPPRLCTDNGIMIANNGIEBRLAGLILHIEIGIEPNCPL  
GVDSKSGEASIVPOLKMEI"

## ORIGIN

## Alignment Scores:

Pred. No.: 1.43e-109 Length: 1820  
Score: 1203.00 Matches: 232  
Percent Similarity: 98.33% Conservative: 3  
Best Local Similarity: 97.07% Mismatches: 4  
Query Match: 97.02% Indels: 0  
Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x AX664695 (1-1820)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
Db CTTGTGGCACTTACTTCAAGAGAGTTTCAGATTCTGCTTCTTGGAAAGCTTTGGACATA 730  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
Db GCACAGGTGACATGCTTGCACAGGTGACAGAGACATTTTAAATTAACATCCAGAG 790  
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyLysGlnGlyLys 60  
Db TGCTCCACATGAGTGTGGTGGAAAGCCATAGAACATTTGGCCAAACAGAAATGATGATT 850  
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
Db CATTTTGAATCAAACTCCCTTGCACTGTCATGCTTAATAATGTGATTTCTTTTACTGA 910  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysGlyLysGlnGlyLysGlnGlyLys 100  
Db CTTCAACAGCTTACTGATTAATAATAATGAAACAGGAAACAGAGGATATTGAGAG 970  
QY 911 CTTCAACAGCTTACTGATTAATAATAATGAAACAGGAAACAGAGGATATTGAGAG 970  
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisSerMetAlaCys 120  
Db GGGCAAAATCTGCTTCCAGACAGACATTTGCTGCCACAGTACAGCACAAATGGCATGT 1030  
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
Db CATCTTGTGAAAAGAACATCGGGCTATTCTGTTTGTGAGCAGAGACCTTGTACT 1090  
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaLysSerAsnPheTyrlleArgArg 160  
Db CAAATAATGACAGTACTGCTGTCATCGTGCTGTCAGTACCTTATATCCGACGA 1150  
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180  
Db 1151 GCTCTGAAATTTTAAACAAACAGACAGTGCATTTGTGTGCTCTCCACAGACTA 1210  
QY 181 CysThrAspAsnGlyIleMetIleAlaIlePheGlnGlyIleGlnArgLeuArgAlaGlyLeu 200

|                |  |   |             |
|----------------|--|---|-------------|
| D <sub>b</sub> | 1211   | TGCACTGATATTGCATTCATTCATTCAGTAATGGAAATTTGAAGAACAACGTACGCTGGCTTG   | 1270        |
| Q <sub>y</sub> | 201  | GlyIleLeuHisAspIleGluGlyIleArgTyrGlnPhePolySerProLeuGlyValAsp     | 220         |
| D <sub>b</sub> | 1271   | GCGATTTTACATGCATCATAAGAAAGCCATCCGCTCATGACCACCAATATGCTCTTGGAGTAAAC | 1330        |
| Q <sub>y</sub> | 221  | IleSerLysGluValAlaGlyValAserIleLeuValProGlnLeuLysMetGluIle        | 239         |
| D <sub>b</sub> | 1331   | ATATCAAAGAAGATTGGAGAACCTTCCTAATAAAGTACACCAATTTAAAATGGAGATA        | 1387        |
| RESULT 9       |  |   |             |
| BC058172       |  |   |             |
| LOCUS          | BC058172   | 1844 bp   | mRNA linear |
| DEFINITION     | Mus musculus cdna clone MGC:67870 IMAGE:5012054, complete cds. |   |             |
| ACCESSION      | BC058172   |   |             |
| VERSION        | BC058172.1   | GI:34849663   |             |
| KEYWORDS       | MGC.   |   |             |
| SOURCE         | Mus musculus (house mouse)                                     |   |             |
| ORGANISM       | Mus musculus   |   |             |

REFERENCE  
AUTHORS  
Strausberg, R. L., Feingold, E. A., Grouse, L. H., Derge, J. G., and  
1 (basses 1 to 1844)  
Eumariota, Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

KLAMNIGER, R.D., COLLINS, F.S., WAGNER, L.H., SHENKMEYER, C.M., SCHULER, G.D., ALTSCHULER, S.F., ZEEBERG, B., BUELOW, K.B., SCHAEFER, C.F., BHAT, N.K., HOPKINS, R.F., JORDAN, H., MOORE, T., MAX, S.I., WANG, J., HATSEH, F., DILCHENKO, L., MARUSINA, K., PARTER, A.A., RUBIN, G.M., HONG, L., STEPLATON, M., SOARES, M.B., DONALDO, M.F., CESAVALI, T.L., SCHECHT, T.E., BROWNSTEIN, M.J., UEDIN, T.B., TOSHTAYLIK, S., CARNINCI, P., PRANGE, C., RATA, S.S., LOQUELLANO, N.A., PETERS, G.J., ABRAMSON, R.D., MULLIN, S.J., BOSAK, S.A., MCBWAN, P.J., MCKERNAN, K.J., MALEK, J.A., GUNARATNE, P.H., RICHARDS, S., WOILEY, K.C., HALE, S., GARCIA, A.M., GAY, L.J., HULY, S.W., VALLATON, D.K., MAZNY, D.M., SODERSTROM, E.J., LU, X., GIBBS, R.A., FAHEY, J., HELTON, E., KETTEMAN, M., MADAN, A., RODRIGUES, S., SANCHEZ, A., WHITING, M., MADAN, A., YOUNG, A.C., SHECHENKO, Y., BOUFFARD, G.G., RODRIGUEZ, A.C., GRIMWOOD, J.W., GREEN, E.D., DICKSON, M.C., RODRIGUEZ, A.C., GRIMWOOD, J., SCHMUTZ, J., MYERS, R.M., BUTERFIELD, Y.S., KRZYWINSKI, M.I., SKALKA, U., SMILLES, D.E., SCHNECH, A., SCHEN, J.E., JONES, S.J. and MARRIS, M.A.,  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
*Proc. Natl. Acad. Sci. U.S.A.* 99 (26), 16899-16903 (2002)

REMARK  
COMMENT

22308257  
MEDLINE  
PUBMED  
12477932  
2 (bases 1 to 1844)  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Strausberg, R.  
Direct Submission  
Submitted (15-SEP-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.C.E. Consortium (LNL)  
DNA Sequencing By: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc.mcgc@nigr.nih.gov](mailto:nisc.mcgc@nigr.nih.gov)  
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
Blakeslee, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brookes, S.,  
Dietrich, N.L., Grant, S., Guan, X., Gupta, J., Haghighi, P.,  
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R.,  
Maduro, O.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
McGowell, L., Pearson, K., Scantripop, S., Thomas, P.J., Touchman, J.W.,  
Tsongen, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
Young, A., Zhang, L.-H. and Green, E.D.

FEATURES  
SOURCE  
1..18144 mscs\_mscs3.tst

**CDS**

misc\_feature

## ORIGIN

|                        |               |
|------------------------|---------------|
| Alignment Scores:      |               |
| Pred. No.:             | 6.02e-98      |
| Score:                 | 1086.00       |
| Percent Similarity:    | 92.05%        |
| Best Local Similarity: | 85.77%        |
| Query Match:           | 87.58%        |
| DB:                    | 10            |
|                        | Gaps:         |
|                        | 0             |
|                        | Length:       |
|                        | 1844          |
|                        | Matches:      |
|                        | 205           |
|                        | Conservative: |
|                        | 15            |
|                        | Mismatches:   |
|                        | 19            |
|                        | Indels:       |
|                        | 0             |

US-10-649-273-2\_COPY\_176\_414 (1-239) x BC058172 (1-1844)

|    |      |  |      |
|----|------|--|------|
| Oy | 1    | LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspLe  | 20   |
| Dd | 747  | CTGTGGCATTAGTCCAAAGTGTTCCGATTTCTGCTCTTGGGAAGCTTTTGACATA      | 806  |
| Oy | 21   | AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuLysHisProGlu    | 40   |
| Dd | 807  | GCGCCAGGCGGACAGTCTTGACAAGGTGGCAAGAAGACTTTTATATCAACATCCAGA    | 866  |
| Oy | 41   | CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe | 60   |
| Dd | 867  | TGTTCTCAATAGAGTGGTGAAGAAAGCTATAGAACATTTGGCCAAAGCGAAATAGATTTC | 926  |
| Oy | 61   | HisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly | 80   |
| Dd | 927  | CATTTACTATCAATCCACCTATGACAGAACTAAGATTTCGATTTTCTTTAATCGGA     | 986  |
| Oy | 81   | LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys | 100  |
| Dd | 987  | CTTCAACATATTACTAGATAGACTATATTAACAACAAGAGAAAGAAAGAGCATTTGAGAG | 1046 |
| Oy | 101  | GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisCysThrCys    | 120  |
| Dd | 1047 | GGGCAATCTCTCATACGCTGCAAGACTTCTGCTGGGTGATACAGCAATCGAAGCGTGC   | 1106 |
| Oy | 121  | HisLeuValLysArgThrHisArgGlnAlaIleLeuPheCysLysGlnArgAspLeuPro | 140  |

| Db         | Accession  | Version | Keywords | Source | Organism | Reference | Authors |
|------------|--|---------|----------|--------|----------|-----------|---------|
| Db         | 1107   |         |          |        |          |           |         |
| Qy         | 141  |         |          |        |          |           |         |
| Db         | 1167   |         |          |        |          |           |         |
| Qy         | 161  |         |          |        |          |           |         |
| Db         | 1227   |         |          |        |          |           |         |
| Qy         | 181  |         |          |        |          |           |         |
| Db         | 1287   |         |          |        |          |           |         |
| Qy         | 201  |         |          |        |          |           |         |
| Db         | 1347   |         |          |        |          |           |         |
| Qy         | 221  |         |          |        |          |           |         |
| Db         | 1407   |         |          |        |          |           |         |
| RESULT 10  |  |         |          |        |          |           |         |
| LOCUS      | BC038910   |         |          |        |          |           |         |
| DEFINITION | Mus musculus O-sialoglycoprotein endopeptidase-like 1, mRNA (CDNA                |         |          |        |          |           |         |
| ACCESSION  | BC038910   |         |          |        |          |           |         |
| VERSION    | BC038910.1   |         |          |        |          |           |         |
| KEYWORDS   | GI:24433548  |         |          |        |          |           |         |
| SOURCE     | Mus musculus (house mouse)   |         |          |        |          |           |         |
| ORGANISM   | Mus musculus   |         |          |        |          |           |         |
| REFERENCE  | Mus musculus   |         |          |        |          |           |         |
| AUTHORS    | 1 (bases 1 to 1017)  |         |          |        |          |           |         |
|            | Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,                     |         |          |        |          |           |         |
|            | Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Sculter, G.D.,         |         |          |        |          |           |         |
|            | Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,           |         |          |        |          |           |         |
|            | Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,            |         |          |        |          |           |         |
|            | Diatchenko, L., Marsina, K., Farmer, A.A., Rubin, G.M., Hong, L.,                |         |          |        |          |           |         |
|            | Capleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,                       |         |          |        |          |           |         |
|            | Scheetz, T.E., Brownstein, M.J., Uedin, T.B., Topolynski, S.,                    |         |          |        |          |           |         |
|            | Carninci, P., Prange, C., Kaha, S.S., Loquellano, N.A., Peters, G.J.,            |         |          |        |          |           |         |
|            | Adkinson, R.D., Mullaly, S.J., Bosak, S.A., McMan, P.J.,                         |         |          |        |          |           |         |
|            | Mckernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,                      |         |          |        |          |           |         |
|            | Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huly, S.W.,                     |         |          |        |          |           |         |
|            | Vallation, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,              |         |          |        |          |           |         |
|            | Fahney, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,                 |         |          |        |          |           |         |
|            | Bancierz, A., Whiting, M., Madan, A., Young, J.W., Shenchenko, Y.,               |         |          |        |          |           |         |
|            | Bouffard, G.G., Blakesley, R.W., Touchman, J.C., Green, E.D.,                    |         |          |        |          |           |         |
|            | Dickson, M.C., Rodriguez, A.C., Greenwood, J., Schmitt, J., Myers, R.M.,         |         |          |        |          |           |         |
|            | Butterfield, Y.S., Krzywinski, M.I., Skalka, J., Smalls, D.E.,                   |         |          |        |          |           |         |
|            | Schmurch, A., Schein, J.E., Jones, S.J. and Marra, M.A.                          |         |          |        |          |           |         |
|            | Generation and initial analysis of more than 15,000 full-length                  |         |          |        |          |           |         |
|            | human and mouse cDNA sequences   |         |          |        |          |           |         |
|            | Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)                        |         |          |        |          |           |         |
|            | 22388257.  |         |          |        |          |           |         |
|            | 12477932   |         |          |        |          |           |         |
|            | 2 (bases 1 to 1017)  |         |          |        |          |           |         |
|            | Strausberg, R.   |         |          |        |          |           |         |
|            | Direct Submission  |         |          |        |          |           |         |
|            | Submitted (25-OCT-2002) National Institutes of Health, Mammalian                 |         |          |        |          |           |         |
|            | Gene Collection (MGC), Cancer Genomics Office, National Cancer                   |         |          |        |          |           |         |
|            | Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,                 |         |          |        |          |           |         |
|            | USA  |         |          |        |          |           |         |
|            | NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a> |         |          |        |          |           |         |
|            | Contact: MGC help desk   |         |          |        |          |           |         |
|            | Email: <a href="mailto:gcgabs-rcmail.nih.gov">gcgabs-rcmail.nih.gov</a>          |         |          |        |          |           |         |
|            | Tissue Procurement: Jeffrey R. Green, M.D.                                       |         |          |        |          |           |         |
|            | cDNA Library Preparation: Life Technologies, Inc.                                |         |          |        |          |           |         |
|            | cDNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LNL)                      |         |          |        |          |           |         |
|            | DNA Sequencing by: Genome Science Centre,  |         |          |        |          |           |         |

```

BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bessel, Yaron Buterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Letlicia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prahnu, Parvaneh Saeedi, Jacqueline
Schrein, Duane Smalins, Michael Smith, Lorraine Spence, Jeff Stoltz,
Michael Thorne, Miranda Tsai, Nataja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LINL at: http://image.llnl.gov
Series: IRAX File: 86 Row: 5 Column: 12.

Location/Qualifiers

1. .1017
   /organism="Mus musculus"
   /mol_type="mRNA"
   /strain="FVB/N"
   /db_xref="taxon:10090"
   /clone="IMAGE:503559"
   /tissue_type="Liver, normal, 5 month old male mouse."
   /clone_id="NCI CGAP L19"
   /lab_host="DH10B"
   /note="Vector: pCMV-SPORT6"

ALIGNMENT SCORES:

Pred. No.:      1.48e-97      Length:      1017
Score:          1079.00      Matches:     205
Percent Similarity: 91.63%      Conservative: 14
Best Local Similarity: 85.77%      Mismatches:  20
Query Match:    87.02%      Indels:      0
BI:              10          Gaps:         0

```

US-10-649-273-2\_COPY\_176\_414 (1-239) x BC038910 (1-1017)  
 QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyIysSerLeuAspIle 20  
 Db 110 CTGTTGGCATTAAGTCCAAAGTGTTCCTCCGATTTCTCGTCTCTGGGAAGCTTTTGGACATA 169  
 QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleIysHisProGlu 40  
 Db 170 GCACCAAGGAGCATCTTGCACAAAGTGGCAAGAGACCTTTCTTTAATCAACATCCAGAA 229  
 QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaIysGlnGlyAsnArgPhe 60  
 Db 230 TGTTCATCATGACGTGGTGGAAGAACTATAGAACAGTTGGCCAAAGACGAATAGATTTC 289  
 QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
 Db 290 CATTTTACATCAATCCACCTTATGAGAAATGCTAAGAAATTGGCATTTTCTTTCACGGGA 349  
 QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIleGluLys 100  
 Db 350 CTTCAACATATTAAGTATAGCTTAATACACACAAAGAAAAGAAAGACATTGGAAG 409  
 QY 101 GlyGlnIleLeuSerSerIleAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 120  
 Db 410 GGGCAAAATCTGTCACTCAGCTCCAGACATTTGTCTCGCGTATACAGATGCACACGGCTGC 469  
 QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140  
 Db 470 CACCTTGGCAAAAAGAACACATGCCGTATTTCTGTTTGGACAGAGAAAATTTGCTTCT 529  
 QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspNhePheTyrIleArgArg 160  
 Db 530 CCAGCTTAACGCGATTTAGTTGTAATCTGGAGGTTTGCAGATGACTTGTATCATCCGAAA 589  
 QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
 Db 590 GCATTGGAAATTTGTGCAAAATCCAAACGACGTCACTGTGTGTGTCCACCTCCAAAGCTG 649



|  |  |  |                                  |        |        |
|--|--|--|----------------------------------|--------|--------|
| QY   | 181  | Cythrhaplangi.Y11  | emeC11a12rpsnsg1y1leg1ua3ylea3A1 | AG1yeu | 200    |
| Db   | 650  | TGCACGTACAATGGCATCATCATATGTCAGATGGAATGGAAAGATTACGTCGTGGCTTG  |                                  |        | 709    |
| QY   | 201  | G1yle1LeuH1sAp11eG1uG1Y11e                                   | grgYrG1uProlYsCySProLeuG1Yva1Asp | 220    |        |
| Db   | 710  | GGCGTTTACATGATGTAGAAAGACATCCGATATGACCAAAATGTCTCTTGAGTAGAC    |                                  |        | 769    |
| QY   | 221  | 11eSerYsg1uVa1G1Yg1uA1sSer11eYsVa1ProG1nLeuYsmeCg1u1e        | 239                              |        |        |
| Db   | 770  | ATATCCAGAGAAATGGCAGAAAGCTGCCATAAAGTACCGCATTTAAAAATGGCACTT    | 826                              |        |        |
| RESULT 11  |  |  |                                  |        |        |
| LOCUS  | AX713716   |  | 2208 bp                          | DNA    | linear |
| DEFINITION   | Sequence 400 from Patent EP1293569.  |  |                                  |        |        |
| ACCESSION  | AX713716   |  |                                  |        |        |
| VERSION  | AX713716.1   | GI:29888642  |                                  |        |        |
| KEYWORDS   |  |  |                                  |        |        |
| SOURCE   |  |  |                                  |        |        |
| ORGANISM   | Homo sapiens (human)   |  |                                  |        |        |
| REFERENCE  |  |  |                                  |        |        |
| AUTHORS  | 1 Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Tanecikita, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and Masuho, Y. |  |                                  |        |        |
| TITLE  | Full-length cDNAs  |  |                                  |        |        |
| JOURNAL  | Patent: EP 1293569-A 400 19-MAR-2003;  |  |                                  |        |        |
| FEATURES   | Helix Research Institute (JP) ; Research Association for Biotechnology (JP)  |  |                                  |        |        |
| SOURCE   | location/Qualifiers  |  |                                  |        |        |
|  | 1..2208  |  |                                  |        |        |
|  | /organism="Homo sapiens"   |  |                                  |        |        |
|  | /mol_type="unassigned DNA"   |  |                                  |        |        |
|  | /db_xref="taxon:9606"  |  |                                  |        |        |
| ORIGIN   |  |  |                                  |        |        |
| Alignment Scores:  |  |  |                                  |        |        |
| Pred. No.:   | 3.57e-95   | Length:  | 2208                             |        |        |
| Score:   | 1059.00  | Matches:   | 211                              |        |        |
| Percent Similarity:                                      | 88.28%   | Conservative:  | 0                                |        |        |
| Best Local Similarity:                                   | 88.28%   | Mismatches:  | 4                                |        |        |
| Query Match:   | 85.40%   | Indels:  | 24                               |        |        |
| DB:  | 6  | Gaps:  | 1                                |        |        |
| US-10-649-273-2_COPY_176_414 (1-239) x AX713716 (1-2208) |  |  |                                  |        |        |
| QY   | 1  | 1Leu1eua1a1eua1G1nG1Yva1sSerApPheLeu1eua1G1YvysSer1eua1p11e  | 20                               |        |        |
| Db   | 869  | CTGTGGCATTAAGTTCACAGAGATTTCACATTTTCGCTTCGGAAGATCTTTTGACATA   | 928                              |        |        |
| QY   | 21   | AlApProG1YaSPMe1LeuAsp1yVa1AlaArgArG1eUsSer1eU11eYsH1sProG1u | 40                               |        |        |
| Db   | 929  | GCACCAAGGTGACATCTTCGACAGGTGGCAGAGACATTCCTTTAATTAACATCCAGG    | 988                              |        |        |
| QY   | 41   | CysSerThm1eSerG1Yg1Yb1sAla11eG1uH1sLeua1a1yG1nG1Yva1snaRPh   | 60                               |        |        |
| Db   | 989  | TGCTCCACATGAGTGGTGGGAAACCATTAAGACATTTGGCCAAACAAAGAAATGATTT   | 1046                             |        |        |
| QY   | 61   | H1sPheAsp11eYsProProl1eH1sH1sAlaYsAsnCySAsPheSerPheThrG1Y    | 80                               |        |        |
| Db   | 1049   | CATTTCACATCAACCTCCCTTCATCATGCTTAATAAATTTGATTTCTTTTTCCTGGA    | 1108                             |        |        |
| QY   | 81   | LeuG1nH1sVa1ThAsp1yG11e11eMe1y1sG1uYsG1uG1Y11eG1u1yS         | 100                              |        |        |
| Db   | 1109   | CTTCAACACGTACATGATTAATAATATGAAAAAGAAAAAGAGAAAGTATTGAGAG      | 1166                             |        |        |
| QY   | 101  | G1Yg1n11eLeuSerSerAla1a1Asp11eAla1a1rVa1G1nH1sThm1eAlaCyS    | 120                              |        |        |
| Db   | 1169   | GGGCAAAATCTCTTTCAGCAGACATGCTGGCAAGTACAGACACAAATGGCATGT       | 1222                             |        |        |

| Accession | Version | Keywords    | Source                                     | Organism     | Reference | Authors   |
|-----------|---------|-------------|--|--------------|-----------|---|
| AK055441  | 1       | GI:16550166 | oligo capping; fis (full insert sequence). | Homo sapiens | 1         | Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Setine, M., Oyamato, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagaishi, K., Murakami, K., Yasuda, T., Iwayanagi, T., Magetsuna, M., Shiratori, A., Sudo, H., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Niimiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yoshida, M., Hoshida, T., Kusano, J., Kanehori, K., Takahashi, Fujii, A., Hara, T., Tanase, T., Nomura, Y., Togiy, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Mutsashino, K., Yuuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohara, N., Sano, S., Moriya, S., Momiyama, H., Satch, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kunagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hiraoka, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Ohtani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuh, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. |
| AK055441  | 1       | GI:16550166 | oligo capping; fis (full insert sequence). | Homo sapiens | 1         | Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Setine, M., Oyamato, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagaishi, K., Murakami, K., Yasuda, T., Iwayanagi, T., Magetsuna, M., Shiratori, A., Sudo, H., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Niimiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yoshida, M., Hoshida, T., Kusano, J., Kanehori, K., Takahashi, Fujii, A., Hara, T., Tanase, T., Nomura, Y., Togiy, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Mutsashino, K., Yuuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohara, N., Sano, S., Moriya, S., Momiyama, H., Satch, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kunagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hiraoka, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Ohtani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuh, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. |
| AK055441  | 1       | GI:16550166 | oligo capping; fis (full insert sequence). | Homo sapiens | 1         | Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Setine, M., Oyamato, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagaishi, K., Murakami, K., Yasuda, T., Iwayanagi, T., Magetsuna, M., Shiratori, A., Sudo, H., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Niimiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yoshida, M., Hoshida, T., Kusano, J., Kanehori, K., Takahashi, Fujii, A., Hara, T., Tanase, T., Nomura, Y., Togiy, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Mutsashino, K., Yuuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohara, N., Sano, S., Moriya, S., Momiyama, H., Satch, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kunagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hiraoka, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Ohtani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuh, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. |
| AK055441  | 1       | GI:16550166 | oligo capping; fis (full insert sequence). | Homo sapiens | 1         | Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Setine, M., Oyamato, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagaishi, K., Murakami, K., Yasuda, T., Iwayanagi, T., Magetsuna, M., Shiratori, A., Sudo, H., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Niimiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yoshida, M., Hoshida, T., Kusano, J., Kanehori, K., Takahashi, Fujii, A., Hara, T., Tanase, T., Nomura, Y., Togiy, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Mutsashino, K., Yuuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohara, N., Sano, S., Moriya, S., Momiyama, H., Satch, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kunagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hiraoka, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Ohtani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuh, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. |
| AK055441  | 1       | GI:16550166 | oligo capping; fis (full insert sequence). | Homo sapiens | 1         | Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Setine, M., Oyamato, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagaishi, K., Murakami, K., Yasuda, T., Iway   |



Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsumoto, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Waga, T., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuno, Y., Nagai, K. and Isogai, T., NEDO human cDNA sequencing project

Unpublished  
3 (bases 1 to 2208)  
Isogai, T., Otsuki, T. and Sugiyama, T.  
Direct Submission  
Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 252-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB) (supported by Japan Key Technology Center etc.); 5' - 3' - end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

FEATURES  
source  
Location/Qualifiers  
1..2208  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="FEBRA2004592"  
/tissue\_type="brain"  
/clone\_id="FEBRA2"  
/dev\_stage="fetus"  
/note="cloning vector: pMB18SFL3"  
344..1438  
/note="unannotated protein product"  
/protein\_id="BAB70923.1"  
/db\_xref="GI:16550167"  
/translation="MLITKTAAGVFFPKSRKRVYEFKRNFGTLFLKIVAGIETSCDDTAAVVDERTGVGEAHSOTVEHLVTGTGIVPAQOOLHRENTORIYOEALSKSVSPDLAATITIKPGLALSLGVLSFSLQVGLQKKPIPHHMRALHTRTNKTVERPVLVLSGGHCLALVQVSDPLILKSLDIADGMDLKYARLPILKPECTMSGKAIENHAKQNNPFDIKPPLHAKNCDPPGLQVHTDKIMKKKESEIEGQILSSADIAATVQHTMAHLVKRTTHRAIIFCQRDLIPQNNAVLVASGVASNFCTIRRALEILTNATQCTILCPPLCTDNGIMIA"

ORIGIN  
Alignment Scores:  
Pred. No.: 3,57e-95 Length: 2208  
Score: 1059.00 Matches: 211  
Percent Similarity: 88.28% Conservative: 0  
Best Local Similarity: 88.28% Mismatches: 4  
Query Match: 85.40% Indels: 24  
DB: 9 Gaps: 1

US-10-649-273-2\_COPY\_176\_414 (1-239) x AK055441 (1-2208)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyValSerLeuAspIle 20  
DB 869 CTTTGGCATTAGTTCAGGAGGTTTCAGATTTCTGCTTGGAGAGCTTTCGACACTA 928  
QY 21 AlaProGlyAspMetLeuAspIleValAlaArgArgLeuSerLeuIleValSerProGlu 40  
DB 929 GACACGAGGTACATGCTTGCACAGAGTGCAGAGACTTCTTAATTAACATCCAGAG 988  
QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaValGlnGlyValAsnArgPhe 60  
DB 989 TCTTCCACCATGAGTGGTGGAGAAAGCATAGAACATTGGCCAAACAGAAATAGATT 1048  
QY 61 HisPheAspIleValSerProLeuHisAlaAlaValAsnGlyAspPheSerPheThrGly 80  
DB 1049 CATTTCGACATCAAACTCCCTTCGACATCAATGTAATTTGATTTCTTTCCTTACTGGA 1108  
QY 81 LeuGlnHisValThrAspIleIleMetLysGlyGluValGlnGlyValGlnGlyLys 100

DB 1109 CTTCAACAGCTTACTGATTAATTAATGAAAAAGAGAAAGATTTGACAGAG 1168  
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120  
DB 1169 GGGCAATTCCTGCTTTCAGCAGCAGACATCTCTGCAAGTACAGACATGACATGT 1228  
QY 121 HisLeuValIysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
DB 1229 CATCTTGAAAAAGAACATCGGGCTTCTGTTTGTAAAGACAGAGACTGTTTACCT 1288  
QY 141 GlnSerMetAlaValLeuValAlaSerGlyValAlaSerAspPheTyrIleArgArg 160  
DB 1289 CAATATATATGAGATCGTGTGATCGTGTGTGCGAAGTAACTTGTATCCGACGA 1348  
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
DB 1349 GCTTCGAAATTTTACCAAGCAGACAGTGCATTTGTGTGTCTCTCCGACACTA 1408  
QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
DB 1409 TGCACTGATATATGACATATGATTCGA----- 1435  
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220  
DB 1436 -----TGATGTCTCTTGGAGTACAC 1456  
QY 221 IleSerLysGluValGlyGluAlaSerIleLysValProGlnIleLeuMetGlnIle 239  
DB 1457 ATATCAAAAAGAGTTGGAGAGCTTCATTAAGTACACAAATTAAATGAGATTA 1513

RESULT 13  
BC078974  
LOCUS  
DEFINITION  
Rattus norvegicus cDNA clone IMAGE:7111906, partial cds.  
ACCESSION  
BC078974.1 GI:50926879  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 1546)  
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
Diatchenko, L., Marusina, K., Palmer, A.A., Rubin, G.M., Hong, L.,  
Stadleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Schaefer, T.E., Brownstein, M.J., Udell, T.B., Toshiyuki, S.,  
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullany, S.D., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huijy, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahy, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Scherchenko, Y.,  
Bonfard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E.,  
Schercher, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 1546)  
Director MGC Project.  
DIRECT SUBMISSION  
Submitted (02-AUG-2004) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>

## COMMENT

Contact: MGC help desk  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Howard Jacobs  
 CDNA Library Preparation: Express Genomics  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
 Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www-shgc.stanford.edu>  
 Contact: (Dickson, Mark) mcdpaxil.stanford.edu  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
 R. M.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
 Series: IRAC Plate: 182 Row: f Column: 6  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Hexamer frequency ORF  
 analysis, Similarity but not identity to protein.  
 Location/Qualifiers

## FEATURES

## CDS

```

1..1546
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:7111906"
/tissue_type="Testis, rat (Brown Norway)"
/clone_lib="NIH MGC_237"
/lab_host="DH10B"
/notes="Vector: pExpress1"
144..>1546
/codon_start=1
/product="unknown (protein for IMAGE:7111906)"
/protein_id="AAH78974.1"
/db_xref="GI:50926880"
/translation="MMLSTKAGAIIPPPSNVGFIRFNVQRALPHNLVGIET
SCDDYAAVDETNVGEALHSQTEVHLTKGI VPPAQOQHEENQRTVEALASAS
GVSPDLSAATTKPGIALSGVLSVQVLPVQPKPFIHMEHALITLTK
VGFPSVLISGHCILALVQSVSDPLFGSLDIAPDMDKVARSLIKHECST
MSGKAIENLAKENRPHFTINPMONAKNDPSPTGQHTDILITHEKEGIEKG
QIISSAADIAAIVHATACHAKRTHAIIIFCOCKNLSPPANAVLVSGVSNLYIR
RALEIVANATQCTILCPPRLCTDNGIMIANGLERLAKAGLILHVEDIRYEPKAGE
IVSMRLVLTALTELDLSVSTHTVAHSPLSNGRANQIOTMCSCTQTVMYRVTNLT
NINERKSKKKKKKKKK"

```

## ORIGIN

## Alignment Scores:

Pred. No.: 8.27e-88 Length: 1546  
 Score: 983.00 Matches: 188  
 Percent Similarity: 92.06% Conservative: 9  
 Best Local Similarity: 87.85% Mismatches: 17  
 Query Match: 79.27% Indels: 0  
 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x BC078974 (1-1546)

```

QY 1 LeuleuAlaleuValGInglyValSerAspPheLeuLeuGlyLyseSerLeuAspIle 20
DB 669 CTGTGGCGGTAAAGTGGAGAGTTCAGATTTTCCTCGCGAAGTCCCTGAGCAATA 728
QY 21 AlaprodIlyaspMetLeuAspLyseValAlaIargArgLeuSerLeuIlyLeHisProGlu 40
DB 729 GCGCCAGGCGACGCTTGACAGAGTGGCAGAGACCTTCTTTAATCAACATCCAGAA 788
QY 41 CysSerTherMetSerGlyLyseAlaIleGluHisLeuAlaIysGInglyAsnArgPhe 60
DB 789 TGTTCATCAATGAGTGGGAGAAAGCTATAGAACCTTTGGCCAAAGAAAGAAATGATTC 848
QY 61 HisPheAspIlyLeysProProLeuHisIleSalAlaIysAsnGlyAspPheSerPheThrGly 80
DB 849 CACTTACTACTCAATCCACCCAGATGCAATGAACTGATTTTCCTTTTACGGGGA 908
QY 81 LeuGlnHisValIThrAspLyseIleIleMetLyseGlyGluGlyGluGlyIleGluLys 100
DB 909 CTTCACATGTCACCGAATGCTAATTAACACACAGAAAGAAAGGACATGAGAG 968

```

```

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaIthrValGlnHisThrMetAlaCys 120
DB 969 GGGCAATCCCTGTCATCAGCCGACAGACATGCTGCGGTACACAGCAGCAACACGTCGC 1028
QY 121 HisLeuValIyethArgThHisArgAlaIleLeuPheCysGlyGlnArgAspLeuPro 140
DB 1029 CACCTTGGAAAGAAACACATCGCTCTTCTGTTTGCACAGCAAAAATTTGCTATCT 1088
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyLyseValAlaIasSerAspPheThrIleArgArg 160
DB 1089 CCAGCTAACGCAATGATTTAGTTGTCTGTCTGAGGTGTTGCAAGTACTTGACATCCGAGA 1148
QY 161 AlaleuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
DB 1149 GCATTGAAATTTGTGCAATGCAACACATGCACTTTGTGTGTCCCTCCGAGACTG 1208
QY 181 CysThrAspAsnGlyIleMetIleAlaIthrAsnGlyIleGluArgLeuArgAlaIleu 200
DB 1209 TGCACTGCAATGTGATCATGATTCATGATGATGAAATTTGAAAGATTACGTCTGCTTG 1268
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrgIuProLys 214
DB 1269 GCATTTCATGATGATGAGAGACATCCGATTCGAAACCAAG 1310

```

## RESULT 14

BC051211

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

## ORGANISM

Mus musculus

(house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 1109)

Strausberg, R.

Submitted (14-APR-2003) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

## REFERENCE

AUTHORS

TITLE

JOURNAL

Direct Submission

Submitted (14-APR-2003) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

## REMARK

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>

Contact: MGC help desk

Email: cgabbs-remail.nih.gov

Tissue Procurement: Marcello Bento Soares, Ph.D.

CDNA Library Preparation: M. Bento Soares, University of Iowa

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>

## COMMENT

contact: amadan@systemsbio.org

Anup Madan, Jessica Fahy, Erin Helton, Mark Kettelman, Anuradha

Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>

Series: IRAC Plate: 113 Row: b Column: 1

This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF

## FEATURES

location/Qualifiers

1..1109

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:1327545"

/tissue\_type="Thymus gland, mouse"

## ORIGIN

/clone\_lib="Soares\_thymus\_2bmt"

/lab\_host="DH10B"

/note="Vector: pT7T3D-Pac"

## Alignment Scores:

Pred. No.: 3,72e-84 Length: 1109  
 Score: 944.50 Matches: 180  
 Percent Similarity: 82.52% Conservative: 20  
 Best Local Similarity: 74.39% Mismatches: 32  
 Query Match: 76.17% Indels: 11  
 DB: 10 Gaps: 2

US-10-649-273-2\_COPY\_176\_414 (1-239) x BC051211 (1-1109)

```

QY 2 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 21
DB 148 TTGAAACTGACTGTCATAGTGGGTTATCTTGTCTTAAT-----AATTTGCAG 198
QY 22 ProGlyAspMetLeuAspLys-----ValAlaArgArgLeu 33
DB 199 CCAAAATATATTTATGAAAAAATGTAATGCTTTTAAATAGGTGGCAAGACCTT 258
QY 34 SerLeuIleLysHisAspProGlyCysSerThrMetSerGlyLysAlaIleGluHisLeu 53
DB 259 TCTTTATCAAAACATCCAGAAATGTTCTACATGATGATGGAAGAAAGCTATAGAACCTTG 318
QY 54 AlaValGlnGlyAsnArgPheHisPheAspIleLysProLeuHisHisAlaValAsn 73
DB 319 GCCAAAGACGGAATGATTCATTTTATCATCATCCACTATGACAAATGCTAAGAAAT 378
QY 74 CysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysValGlu 93
DB 379 TGGGATTTTCTTTCACGGGACTTTCACATATATTCATATAGCTAATTAACACACAAAGAA 438
QY 94 LysGlnGluGlyIleGlnLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThr 113
DB 439 AAAGAAAGAGGACTTGAAGAGGGCAATCTGTCTCATCTGACATGACATTTGCTGCGG 498
QY 114 ValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCys 133
DB 499 GTACAGCATGCAACACGCTGCCACTTTCGCAAAAAGAACACATCGCGCTATTTGTTTGC 558
QY 134 LysGlnArgAspLeuLeuProGlnAsnAspAlaValLeuValAlaSerGlyGlyValAla 153
DB 559 AAGCAAAAATTTGCTCTCTCCAGCTTAACGCAAGTATTAAGTTGATGGAAGGTGGCA 618
QY 154 SerAsnPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeu 173
DB 619 AGTAATTTGATCATCCGAAAGCATTTGGAATTTGCGCAATATGCAAGCAGTGCACCTTG 678
QY 174 LeuCysProProArgLeuGlySerThrAspAsnGlyIleMetIleAlaIleAspGlyIle 193
DB 679 TTGTGTCACCTTCACAGACTGTCACATGACATGACATGATGATGCAATGGAAT 738
QY 194 GlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgGlyPro 213
DB 739 GAAAGATTACGCTGCTGCTGGCTGCTTTTACATGATGTAGAACACATCCATATGAAACA 798
QY 214 LysCysProLeuGlyValAspIleSerLysGlnValGlyGlnAlaSerIleLysValPro 233
DB 799 AATATGCTCTTGGAGTGAATTCACATTCACAGAAAGTTGCAAGAGCTGCATTAAGATCCG 858
QY 234 GlnLeuLysMetGluIle 239
DB 859 CGATTAAATATGCACTT 876

RESULT 15
LOCUS BX930694 860 bp mRNA linear VRT 30-MAR-2004
DEFINITION Gallus gallus finished cDNA, clone CHEST967014.
ACCESSION BX930694
VERSION BX930694.2 GI:46016745
KEYWORDS
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

## REFERENCE

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 1 (bases 1 to 860)  
 Boardman, P.E., Bonfield, J.K., Brown, W.R.A., Carder, C., Chalk, S.E.,  
 Croning, M.D.R., Davies, R.M., Francis, M.D., Grafham, D.V.,  
 Hubbard, S.J., Humphrey, S.J., Hunt, P.J., Muddison, M., McLaren, S.R.,  
 Niblett, D., Overton, I.M., Rogers, J., Scott, C.E., Taylor, R.G.,  
 Tickle, C. and Wilson, S.A.  
 Direct Submission

## TITLE

## JOURNAL

Submitted (29-MAR-2004) Sanger Institute, Hinxton, Cambridgeshire,  
 CH10 1SA, UK. E-mail enquiries: chicken@hms.wmst.ac.uk

## COMMENT

On Apr 1, 2004 this sequence version replaced gi:41631222.  
 BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA  
 sequencing project.  
 This sequence is from the  
 BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,  
 from a library constructed by Elizabeth Bosch. cDNA was prepared  
 from RNA extracted from limbs, normalised, and poly A-tailed.  
 EcoRI-NciI cut cDNA was then ligated into the vector. Vector:  
 pBlueScript II KS(+); Site 1: EcoRI; Site 2: NciI Host: Escherichia  
 coli DH10B.

## FEATURES

## source

1..860  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="White Leghorn, Hisex"  
 /db\_xref="taxon:9031"  
 /clone="CHEST967014"  
 /clone\_1ib="CS90CHN59"  
 /dev\_stage="stage 36"

## ORIGIN

## Alignment Scores:

Pred. No.: 7.86e-71 Length: 860  
 Score: 809.00 Matches: 151  
 Percent Similarity: 79.66% Conservative: 37  
 Best Local Similarity: 63.98% Mismatches: 48  
 Query Match: 65.24% Indels: 0  
 DB: 5 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x BX930694 (1-860)

```

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB 94 ATCTTGGACAGTACAGACAGAGATTTCAGATTTCTTCTGCTGACATCATATGATTA 153
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 154 GCACACAGTGCATGATGATTAAGTGAAGAGGCTCTTTAAGTGAACACACCGGAG 213
QY 41 CysSerThrMetSerGlyLysValAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 214 TGCACAGCATGCGCGGGGGAAGCAATAGACCTGCTCAAAACCGAGACTGGCAA 273
QY 61 HisPheAspIleLysProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80
DB 274 CAGTACACTTTCACACTTCCATGACACAGTATGTAATCTGATTTCTTCTCCGGA 333
QY 81 LeuGlnHisValThrAspLysIleIleMetLysGlnGlyGlnGlyIleGlnLys 100
DB 334 CTTGACAGCTTGTGCACAAAGCCATTTCTGAGAAAGAAAGAAAGATATTCAGAA 393
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 120
DB 394 GGGGAAATCCGTCTTCGTTAAGACATGCTCTGTCACAGCAGATGATGCTGCT 453
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB 454 CATATTTCCAGCGGACACACCGACCATGCTCTTCTGCAATGAAAACACATATTA 513
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaAspAsnPheThrIleArgArg 160
DB 514 CCAAAAAGTCACTGCTTGTGTATCAGAGGAGTTCAGATTAATCATATATGAGAAA 573

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: June 16, 2005, 12:03:59 ; Search time 345.511 Seconds  
(without alignments)  
4094.859 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414  
Perfect score: 1240  
Sequence: 1 LIALVQGVSDFLILGKSLDI.....DISKVGKASIKVPOLNMEI 239

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 4390206 seque, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+g2n.model -DEV=xlp  
-Q=/cgn2.1/USPTO.spool\_p/US10649273/runat\_15062005\_111415\_6022/app\_query.fasta.1.1429  
-DB=N Geneseg 16Dec04 -OFMT=faactap -SUFFIX=g2n.rng -MIMMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10649273\_@CGN\_1.1\_1063\_@runat\_15062005\_111415\_6022 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: N\_Geneseg\_16Dec04:\*

1: genesegq1980s:\*\n2: genesegq1990s:\*\n3: genesegq2000s:\*\n4: genesegq2001as:\*\n5: genesegq2001bs:\*\n6: genesegq2002as:\*\n7: genesegq2002bs:\*\n8: genesegq2003as:\*\n9: genesegq2003bs:\*\n10: genesegq2003cs:\*\n11: genesegq2003ds:\*\n12: genesegq2004as:\*\n13: genesegq2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| 1          | 1240  | 100.0       | 1416   | 8  | ABX70950 Novel hum |
| 2          | 1240  | 100.0       | 1526   | 6  | ABs76639 DNA encod |
| 3          | 1240  | 100.0       | 2058   | 6  | ABs93268 Human O-c |
| 4          | 1240  | 100.0       | 2197   | 6  | ABs76635 DNA encod |
| 5          | 1240  | 100.0       | 2572   | 8  | ABT23207 Human pro |

|    |       |      |       |    |                                |
|----|-------|------|-------|----|--------------------------------|
| 6  | 1203  | 97.0 | 1820  | 6  | AAD46856 Human g1y             |
| 7  | 1203  | 97.0 | 1820  | 6  | ACA60887 Human cDN             |
| 8  | 1203  | 97.0 | 1821  | 10 | ABSS7020 cDNA enco             |
| 9  | 1059  | 85.4 | 2208  | 10 | ADAs2832 Human cod             |
| 10 | 1059  | 85.4 | 2890  | 12 | ADQ24627 Human sof             |
| 11 | 995.5 | 80.3 | 3358  | 10 | ADE31345 Human dia             |
| 12 | 870   | 70.2 | 1572  | 6  | ABQ75508 Murine si             |
| 13 | 599   | 48.3 | 2734  | 5  | AAS84622 DNA encod             |
| 14 | 468   | 37.7 | 371   | 12 | ADL86725 DNA up-re             |
| 15 | 468   | 37.7 | 371   | 12 | ADL86726 DNA up-re             |
| 16 | 332.5 | 26.8 | 1601  | 4  | ABL24633 Drosophil             |
| 17 | 332.5 | 26.8 | 1656  | 4  | ABL24632 Drosophil             |
| 18 | 322   | 26.0 | 1557  | 3  | AAC38454 Arabidops             |
| 19 | 257   | 20.7 | 1146  | 8  | ACA26804 Prokaryot             |
| 20 | 248.5 | 20.0 | 1000  | 4  | AAF91424 Moraxella             |
| 21 | 248.5 | 20.0 | 1000  | 6  | ABK37804 DNA seque             |
| 22 | 248.5 | 20.0 | 1044  | 8  | ACA39102 Prokaryot             |
| 23 | 248.5 | 20.0 | 94750 | 4  | AAF28551 Genomic f             |
| 24 | 247   | 19.9 | 936   | 6  | ACA20445 Prokaryot             |
| 25 | 247   | 19.9 | 1053  | 12 | ADL03120 DNA encod             |
| 26 | 241   | 19.4 | 1032  | 8  | ACA43173 Prokaryot             |
| 27 | 239   | 19.3 | 4360  | 6  | AD48239 Ehrlichia              |
| 28 | 235.5 | 19.0 | 1092  | 6  | ABQ90383 M. capaul             |
| 29 | 234   | 18.9 | 300   | 3  | AAH00934 Human col             |
| 30 | 232   | 18.7 | 676   | 4  | AAH08019 Human CDN             |
| 31 | 232   | 18.7 | 1385  | 4  | AAH15110 Human CDN             |
| 32 | 231   | 18.6 | 1026  | 4  | AAH54064 Pseudomon             |
| 33 | 231   | 18.6 | 1026  | 8  | ACA42146 Prokaryot             |
| 34 | 231   | 18.6 | 1026  | 10 | ADG73341 P aerugin             |
| 35 | 231   | 18.6 | 1059  | 11 | ABD02280 Pseudomon             |
| 36 | 231   | 18.6 | 1026  | 10 | ABD02197 P aerugin             |
| 37 | 229   | 18.5 | 1206  | 10 | ADG73343 P aerugin             |
| 38 | 224   | 18.1 | 1029  | 4  | AAS53309 Haemophil             |
| 39 | 224   | 18.1 | 1029  | 4  | ACA34150 Prokaryot             |
| 40 | 224   | 18.1 | 11000 | 2  | AAH42063_05 Continuation (6 of |
| 41 | 217   | 17.5 | 9667  | 13 | ADT05493 Haemophil             |
| 42 | 217   | 17.5 | 85814 | 13 | ADT05493 Haemophil             |
| 43 | 214   | 17.3 | 1044  | 4  | ABL08591 Drosophil             |
| 44 | 213   | 17.2 | 1014  | 4  | AAS56045 Salmonell             |
| 45 | 213   | 17.2 | 1014  | 8  | ACA51431 Prokaryot             |

## ALIGNMENTS

|          |   |                          |
|----------|---|--------------------------|
| RESULT 1 | ABX70950  | standard; cDNA; 1416 BP. |
| XX       | ABX70950;   |                          |
| AC       | ABX70950;   |                          |
| XX       | ABX70950;   |                          |
| DT       | 05-MAR-2003   | (first entry)            |
| XX       | 05-MAR-2003   | (first entry)            |
| DE       | Novel human cDNA sequence #175.   |                          |
| XX       | Novel human cDNA sequence #175.   |                          |
| KW       | Human; gene; ss; nervous system disorder; peripheral neuropathy;          | Aad46856 Human g1y       |
| KW       | Huntington's disease; amyotrophic lateral sclerosis; haemophilia;         | Ac60887 Human cDN        |
| KW       | neurodegenerative disease; Parkinson's disease; Alzheimer's disease;      | ABSS7020 cDNA enco       |
| KW       | autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;   | ADAs2832 Human cod       |
| KW       | insulin-dependent diabetes mellitus; anaemia; thrombocytopaenia; wound;   | ADQ24627 Human sof       |
| KW       | ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;         | ADE31345 Human dia       |
| KW       | fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;       | ABQ75508 Murine si       |
| KW       | coagulation disorder; cancer; tumour; inflammatory disease; septic shock; | AAS84622 DNA encod       |
| KW       | Crohn's disease; anaphylaxis; proliferation; chemotactic;                 | ADL86725 DNA up-re       |
| KW       | differentiation; stem cell growth factor; hematopoiesis; chemokine;       | ADL86726 DNA up-re       |
| KW       | haemostatic; antiinflammatory; expressed sequence tag; EST.               | ABL24633 Drosophil       |
| OS       | Homo sapiens.   | ABL24632 Drosophil       |
| XX       | Homo sapiens.   | AAC38454 Arabidops       |
| XX       | Homo sapiens.   | ACA26804 Prokaryot       |
| PN       | WO200281731-A2.   | AAF91424 Moraxella       |
| XX       | WO200281731-A2.   | ABK37804 DNA seque       |
| PD       | 17-OCT-2002.  | ACA39102 Prokaryot       |
| XX       | 17-OCT-2002.  | AAF28551 Genomic f       |
| XX       | 29-JAN-2002; 2002WO-US001222.   | ACA20445 Prokaryot       |
| PF       | 29-JAN-2002; 2002WO-US001222.   | ADL03120 DNA encod       |

XX 30-JAN-2001; 2001US-00774528.  
PR (HYSE-) HYSEO INC.  
PA (GOOD/) GOODRICH R. W.  
XX  
XX Tang TY, Liu C, Zhou P, Aundi V, Zhang J, Zhao QA, Ren F,  
PI Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;  
XX WPI; 2003-058563/05.  
DR

PT Novel polypeptide useful for treating neurodegenerative diseases, myeloid  
PT or lymphoid cell disorders, bone disorders, mechanical and traumatic  
PT disorders, coagulation disorders, and inflammatory diseases.  
PS

Claim 1; Page; 612pp; English.

XX This invention relates to the cDNA sequences encoding an isolated novel  
CC human polypeptide. The protein encoded by the nucleic acid of the  
CC invention is useful for treating central and peripheral nervous system  
CC diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic  
CC lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,  
CC Alzheimer's disease); autoimmune disease (e.g. systemic lupus  
CC erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)  
CC ; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopenia)  
CC ; wounds, ulcers, burns; bone disorders (e.g. osteoporosis,  
CC osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head  
CC trauma); lung or liver fibrosis; reperfusion injury in various tissues;  
CC bacterial, viral or fungal infections; allergic conditions such as  
CC allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);  
CC cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's  
CC disease, anaphylaxis). The protein may be used to inhibit the growth,  
CC infection or function of infectious agents such as bacteria, fungi,  
CC viruses, or to effect bodily characteristics, biorhythms or circadian  
CC cycles of rhythms. The protein may also have  
CC proliferation/differentiation, stem cell growth factor, haematopoiesis  
CC regulation, immune stimulation or suppressing, chemotactic/chemokinetic,  
CC haemostatic and thrombolytic, receptor/ligand, and antiinflammatory  
CC activities. The cDNA sequences of the invention are useful for expressing  
CC recombinant protein for analysis. The present sequence represents a novel  
CC human cDNA sequence of the invention, this sequence is an expressed  
CC sequence tag (EST) and was identified using subtractive hybridisation  
XX

Sequence 1416 BP; 441 A; 280 C; 273 G; 422 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 3,21e-132 Length: 1416  
Score: 1240.00 Matches: 239  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 8 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x ABX70950 (1-1416)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyValSerLeuAspIle 20  
DB 586 CTTGTGGCATTAGTTCAGAGAGTTTCAGATTTCCTTCCTTGGAAAGCTCTTGACACTA 645  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
DB 646 GCACCGAGTACATGCTTGCACAGGTGGCAAGAAAGATCTTTCTTAATAAACATCCAGAG 705  
QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
DB 706 TGTCTCCACCATGAGTGGTGGAAAGCCATAGACATTGGCCAAACAGAAATATGATTT 765  
QY 61 HisPheAspLysProProlLeuHisAlaAlaLysAsnGlyAspPheSerPheThrGly 80  
DB 766 CATTTTGACATCAAACTCCCTTCGATCATGCTAAAAATGATGATTTTCTTTTACTGGA 825  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysGlnGlyGlnGlnGlyIleGlnLys 100

DB 826 CTTACACAGCTTACTGATTAATATATGAAAAAGAAAAAGAGAGATTTAGAGAG 885  
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120  
DB 886 GGGCAATCTCTGCTTTCAGCAGCAGCATCTTCTCCACAGACACACATATGGCANT 945  
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
DB 946 CATCTTGAAAAAGAACACATCGGGCTATTCTGTTTGTAAAGACAGAGACTGTACTCT 1005  
QY 141 GlnAsnAsnAlaValIleValAlaSerGlyGlyValAlaSerAsnPheTyrlaArgArg 160  
DB 1006 CAAATATATGACAGTACGTTGTCATCGGTGTCGAGTTCGCAATTACTTTATATCCGAGA 1065  
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
DB 1066 GCTTGGAAATTTTAAACAAACGACACAGGCACTTTGTTGTCTCTCCACAGACTA 1125  
QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
DB 1126 TGCACTGATATAGCATATATGATTCATGCAAGAAATGTAAGAAAGACTACGTCGCTTG 1185  
QY 201 GlyIleLeuHisAspIleGlnGlyIleArgTyrlGluProLysCysProLeuGlyValAsp 220  
DB 1186 GGCATTTTACATGACATAGAAAGCATCCGCTATTAACCAAAATGTCTCTTGGAGTAGAC 1245  
QY 221 IleSerLysGlnValGlyGlnAlaSerIleLysValProGlnLeuLysMetGlnIle 239  
DB 1246 ATATCAAAAGAGATTGGAGAAAGCTTCATTAAGATACCAATTAATAATGAGATA 1302  
RESULT 2  
ABS76639  
ID ABS76639 standard; DNA; 1526 BP.  
XX  
AC ABS76639;  
XX  
DT 11-DEC-2002 (first entry)  
XX  
DE DNA encoding novel human metalloprotease MP1 fragment #1.  
XX  
XX Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;  
KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
KW liver disease; renal disease; immune disorder; rheumatoid arthritis;  
KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
KW neurological disorder; gene; ds.  
OS Homo sapiens.  
XX  
XX WC200272751-A2.  
XX  
XX 19-SEP-2002.  
XX  
XX 05-FEB-2002; 2002W0-US003353.  
XX  
XX 05-FEB-2001; 2001US-0266518P.  
XX  
XX 10-APR-2001; 2001US-0282814P.  
XX  
XX (BRIM) BRISTOL-MYERS SQUIBB CO.  
XX  
XX Chen J, Feder J, Nelson TC, Duclos F, Krystek S;  
XX WPI; 2002-723329/78.  
XX  
XX P-PSDB: ABG96487.  
XX  
XX New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
PT treating, or ameliorating diseases associated with aberrant  
PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and  
PT neurological disorders.

XX Disclosure; Page 462-463; 473pp; English.

CC The invention describes an isolated nucleic acid molecule (I) encoding a  
CC metalloprotease (MP-1). (I) is useful for preventing, treating, or  
CC ameliorating a medical condition, particularly an immune disorder, an  
CC aberrant glutamate transport or motor neuron disorder, such as  
CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
CC condition. The composition and methods are also useful for diagnosing,  
CC prognosticating, treating, ameliorating and/or treating disorders  
CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
CC Alzheimer's disease or Parkinson's disease). This sequence represents a  
CC metalloprotease MPI polynucleotide

XX Sequence 1526 BP; 484 A; 297 C; 289 G; 456 T; 0 U; 0 Other;

#### Alignment Scores:

|                        |           |               |      |
|------------------------|-----------|---------------|------|
| Pred. No.:             | 3,57e-132 | Length:       | 1526 |
| Score:                 | 1240.00   | Matches:      | 239  |
| Percent Similarity:    | 100.00%   | Conservative: | 0    |
| Best Local Similarity: | 100.00%   | Mismatches:   | 0    |
| Query Match:           | 100.00%   | Indels:       | 0    |
| DB:                    | 6         | Gaps:         | 0    |

US-10-649-273-2\_COPY\_176\_414 (1-239) x ABA93268 (1-1526)

```
Qy 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
Db CTGTGGCAATTAAGTTCAGAGAGTTTCAGATTCTGCTTGGAAAGCTTTGGACATA 144
Qy 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
Db 145 GCACCAAGTGTACATGCTTGAACAAGTGGCAAGAACCTTCTTAAATAAACATCCAGAG 204
Qy 41 CysSerThrMetSerGlyLysValAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
Db 205 TGGTCCACCAATGAGTGGGAAAGCCATACACATTTGGCCAAACAGAAATAGATT 264
Qy 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
Db 265 CATTTGACATCAAACTCCCTTCATCATCTTAAATTTGATTTCTTTTACTGGA 324
Qy 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIleGluLys 100
Db 325 CTTCAACACCTTACTGATTAATTAATGAAAAAGGAAAAAGGAGATTGAGAG 384
Qy 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaThrValGlnHisThrMetAlaCys 120
Db 385 GGGCAATCTGCTTTCAGACGACGACATTCCTCCACAGTACACACATGCGCATGT 444
Qy 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
Db 445 CATCTTGTGAAAGACACATCGGGCTATTCTGTTTGTAAAGCAGAGACTTGTACT 504
Qy 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspPheTyrIleArgArg 160
Db 505 CAATAATATGACAGTACGTTGATCGTGTGCGCAATTAATCTTATATCGCAGA 564
Qy 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
Db 565 GCTCTGAAATTTTAACAAACGCAACAGTGCATTTGTGTGTCTCCCTCCACGACTA 624
Qy 181 CysThrAspArgGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db 625 TGACATGATATGCAATTAATGATGCAATGAGATGATTAAGACTACGTCGCTTGG 684
```

```
Qy 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220
Db 685 GGCATTTTACATGACATAGAGGATCGCTATGAAACAAATGTCCTTGGAGTAC 744
Qy 221 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
Db 745 ATATCAAAAAGAGTTGGAGAGCTTCATTAAGTACACATTTAAATGAGATA 801
```

#### RESULT 3

ABA93268 standard; cDNA; 2058 BP.

AC ABA93268;

DT 19-APR-2002 (first entry)

DE Human O-sialoglycoproteinase-like protein encoding cDNA SEQ ID NO:1.

XX Human; O-sialoglycoproteinase-like protein; OSGPLP; enzyme; gene; ss.

XX Homo sapiens.

OS Homo sapiens.

FT Key Location/Qualifiers

FT CDS 110..1354

FT FT /\*tag= a

FT FT /product= "O-sialoglycoproteinase-like protein"

PD CNJ138550-A.

PD 24-OCT-2001.

PF 19-APR-2000; 2000CN-00106834.

PR 19-APR-2000; 2000CN-00106834.

PA (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.

PI Mao Y, Xie Y,

DR WPI; 2002-115090/16.

DR P-PSDB; ABB05481.

XX O-sialoglycoproteinase-like protein and encoding polynucleotide, useful

PT for diagnosing, preventing and treating related diseases.

XX Claim 5; Page 29-30 (Disclosure); 38pp; Chinese.

PS The present sequence encodes human O-sialoglycoproteinase-like protein

CC (OSGPLP). The present invention also describes: (1) the preparation of

CC the OSGPLP protein; (2) applying the OSGPLP protein in diagnosis; (3) the

CC prevention and/or treatment of related diseases; (4) utilizing the OSGPLP

CC protein in screening its agonist, excitomotor and inhibitor and preparing

CC an antibody against the OSGPLP protein; and (5) the use of the OSGPLP

CC polynucleotide sequences, proteins, agonists, excitomotors, inhibitors

CC and antibodies in treating diseases related to the abnormal OSGPLP gene

XX and in preparing the medicine composite for the treatment

XX Sequence 2058 BP; 637 A; 400 C; 410 G; 611 T; 0 U; 0 Other;

#### Alignment Scores:

|                        |           |               |      |
|------------------------|-----------|---------------|------|
| Pred. No.:             | 5,45e-132 | Length:       | 2058 |
| Score:                 | 1240.00   | Matches:      | 239  |
| Percent Similarity:    | 100.00%   | Conservative: | 0    |
| Best Local Similarity: | 100.00%   | Mismatches:   | 0    |
| Query Match:           | 100.00%   | Indels:       | 0    |
| DB:                    | 6         | Gaps:         | 0    |

US-10-649-273-2\_COPY\_176\_414 (1-239) x ABA93268 (1-2058)

```
Qy 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
Db 635 CTGTGGCAATTAAGTTCAGAGAGTTTCAGATTCTGCTTGGAAAGCTTTGGACATA 694
```

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
DB 695 GGCACGAGTGAATGCTTGAACAGGTGGCAAGAGACTTTCTTTATATTAACATCCAGAG 754  
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
DB 755 TGCCTCCACCATGAGTGTGGAGAAAGCCATAGAACATTGGCCAAACAGGAATAGATTT 814  
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
DB 815 CATTGTGACATCAACCTCCCTTGACATCATGCTTAAATATGATTTTCTTTACTGGA 874  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluGlyGlnGlyIleGluLys 100  
DB 875 CTTCAACACGTTACTGATTAATATATGAAAAAGAAAAAGAGAGATTTGAGAG 934  
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120  
DB 935 GGGCAATCCTGCTTCCAGACAGACATTTGCTGCCACATACAGCACATGCGCATGT 994  
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140  
DB 995 CATCTTGAAAAAGAACACATCGCGCTATTCTGTTTGTAGACAGAGACTTGTAACT 1054  
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrlleArgArg 160  
DB 1055 CAAAATATGCACTACTGCTGTGATCTGCTGCTGCTGCCAAGTAACTTATATCCGAGA 1114  
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180  
DB 1115 GCTCTGAAATTTTAAACAAACGACACAGTGCACCTTGTGTGTCCTCCCTCCAGACTA 1174  
QY 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuAlaGlyLeu 200  
DB 1175 TGCACGATATATGATGATTAATGATGATGATTAATGATTAATGATGATGATGATG 1234  
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrlleProLysCysProLeuGlyValAsp 220  
DB 1235 GGCATTTTACATACATAGAGGATCCGCTATGAAACCAAAATGCTCTTGGAGTAGAC 1294  
QY 221 IleSerIleGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239  
DB 1295 ATATCAAAAGAAAGTTGGAGAGAGCTTCATTAAGTATACACATTAATAAATGAGATA 1351

RESULT 4  
ABST76635  
ID ABST76635 standard; DNA; 2197 BP.  
XX  
AC ABST76635;  
XX  
DT 11-DEC-2002 (first entry)  
XX  
DE DNA encoding novel human metalloprotease MPI.  
XX  
KW Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;  
motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
liver disease; renal disease; immune disorder; rheumatoid arthritis;  
acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
neurolysis; cystic fibrosis; vascular disorder; inflammatory disorder;  
neurological disorder; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200272751-A2.  
XX  
PD 19-SEP-2002.  
XX  
PF 05-FEB-2002; 2002MO-US003353.

PR 05-FEB-2001; 2001US-0266518P.  
PR 10-APR-2001; 2001US-0282814P.  
XX  
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S;  
XX  
DR MPI: 2002-723329/78.  
DR P-PSDB; ABG96478.  
XX  
PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
PT treating, or ameliorating diseases associated with aberrant  
PT metalloprotease activity, e.g. immune, metabolic, inflammatory and  
PT neurological disorders.  
XX  
PS Claim 1; Fig 1A-C; 473pp; English.  
XX  
CC The invention describes an isolated nucleic acid molecule (i) encoding a  
CC metalloprotease (MP-1). (i) is useful for preventing, treating, or  
CC ameliorating a medical condition, particularly an immune disorder, an  
CC aberrant glutamate transport or motor neuron disorder, such as  
CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
CC condition. The compositions and methods are also useful for diagnosing,  
CC prognosticating, treating, ameliorating and/or treating disorders  
CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
CC Alzheimer's disease or Parkinson's disease). This sequence represents a  
CC metalloprotease MPI polynucleotide  
XX  
SQ Sequence 2197 BP; 681 A; 441 C; 439 G; 636 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 5.98e-132 Length: 2197  
Score: 1240.00 Matches: 239  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x ABST76635 (1-2197)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
DB 756 CTGTGGCATTAAGTTCAGAGAGTTTCAGATTTTCTGCTTCTTGAAAGTCTTTGGACATA 815  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
DB 816 GCACGAGTGAATGCTTGAACAGGTGGCAAGAGACTTTCTTTATATTAACATCCAGAG 875  
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
DB 876 TGCCTCCACCATGAGTGTGGAGAAAGCCATAGAACATTGGCCAAACAGGAATAGATTT 935  
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
DB 936 CATTGTGACATCAACCTCCCTTGACATCATGCTTAAATATGATTTTCTTTACTGGA 995  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluGlyGlnGlyIleGluLys 100  
DB 996 CTTCAACACGTTACTGATTAATATATGAAAAAGAAAAAGAGAGATTTGAGAG 1055  
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120  
DB 1056 GGGCAATCCTGCTTCCAGACAGACATTTGCTGCCACATACAGCACATGCGCATGT 1115  
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140



|    |  |      |   |      |
|----|--|------|---|------|
| Db |  | 1116 | CATCTTGTGGAAAAGAAACACACTCGGCGTAFTTCGTCTTTTGTAAGCAGAAGACTTGTAACCT    | 1117 |
| Oy |  | 141  | GlnAsnaEnAlAvalleValAlaSerGIylGlYAlAlaSerAsnPhetIrlleaRgaRg         | 160  |
| Db |  | 1176 | CAAAATAATATGCAGTAGCTAGCTGTTGCATCTGCTGCTGCCAAGTAACATTCTAATATCCGAGACA | 1233 |
| Oy |  | 161  | AlAleuglUllEleueThrAsnlArhtGrInCyethrLeuleuCYePrProProkrgleu        | 180  |
| Db |  | 1236 | GCTCTGGAAAAATTTTAACAAACGAACCAACAGTCGACCTTGTTGTCTCTCTCCAGACTTA       | 1293 |
| Oy |  | 181  | CysThraPaSngLIylewettllealatrpaenglyillegLArgleuhRglaglgleu         | 200  |
| Db |  | 1296 | TGCACGTGATATGCATGCGATTTATGTCATGCAAGCATGGAATGATTAAGAACACTAGCTCTGCTTG | 1355 |
| Oy |  | 201  | GLyllleuHiaspIllegluGIyllleaRgyrlUpPolYSCyeproleuGlYvalaSep         | 220  |
| Db |  | 1356 | GGCATTTTTACATGACATGAAGACCATCGCGCTATGAAACCAAATGTCTCTTGAGAGAGAC       | 1411 |
| Oy |  | 221  | IleserlysGIvualGlYgluaLasertlleysValProglntleuLysetheGtulle         | 239  |
| Db |  | 1416 | ATATCCA AAAAGAGATTGGAGAAAGCTTCCTCA TAAAAGTACCACAATTAAAATATGAGATA    | 1472 |

RESULT  
5  
ABTJ23207

|    |   |   |
|----|---|---|
| ID | ABTJ23207   | standard; DNA; 2572 BP.   |
| XX | ABTJ23207;  |   |
| DT | 01-MAY-2003   | (first entry)   |
| DE |   | Human protein modification + maintenance molecule DNA SEQ ID No 36. |
| XX |   |   |
| KW | Cytostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticongulant;<br>cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;<br>antitumor; hepatotropic; gynaecological; antibacterial; viricide;<br>protozoacide; antiparasitic; cell proliferative disease; PMOD;<br>protein modification and maintenance molecule; immunogenic fragment;<br>cancer; autoimmune; inflammatory disease; neurological disorder;<br>gastrointestinal; developmental; vesicle trafficking disorder; infection;<br>protein-protein interaction; drug-target interaction;<br>gene expression profile; human; gene; ds. |   |
| XX | Homo sapiens.   |   |
| OS |   |   |
| XX | WO2003000844-A2.  |   |
| PN |   |   |
| XX | 03-JAN-2003.  |   |
| PD |   |   |
| PF | 18-JUN-2002; 2002MO-USO19360.   |   |
| PR | 22-JUN-2001; 2001US-0300508P.<br>06-JUL-2001; 2001US-0303445P.<br>13-JUL-2001; 2001US-0305405P.<br>09-AUG-2001; 2001US-0311442P.<br>24-AUG-2001; 2001US-0314821P.<br>29-AUG-2001; 2001US-0315992P.<br>03-MAY-2002; 2002US-0378205P.   |   |
| PA | (INCY-) INCYTE GENOMICS INC.  |   |
| XX |   |   |
| P1 | Gandhi AR, Kable AE, Swarnakar A, Hafalia AD, Tran B, Duggan BM,<br>Warren BA, Ison CH, Honcheill CD, Nguyen DB, Lu DAM, Lee EA, Yue H,<br>Foreythe LJ, Barroso I, Ramkumar J, Griffin JA, Li JX, Yang J,<br>Thangavelu K, Gietzen KU, Ding L, Baughm MR, Borowoky MU, Yao MG;<br>Walja NK, Mason PM, Gururajan R, Lee S, Becha SD, Lee SY, Tran UK,<br>Elliott VS, Luo W, Sprague MW, Tang YT, Lu Y, Zebajadian Y;   |   |
| PI | MP1; 2003-184039/18.  |   |
| DR | P-PsDB; ABJ26654.   |   |
| PT | New isolated human PMOD polypeptide and polymucleotide, useful for<br>diagnosing, treating and preventing diseases or conditions associated<br>with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and  |   |

PT infections.

XX Claim 91; Page 211, 225bp; English.

XX The invention relates to an isolated polypeptide comprising: any of 28

CC sequences of 48-1256 amino acids; a natural amino acid sequence at least

CC 90% identical to the 28 amino acid sequences, 94% identical to a sequence

CC of 703 or 267 amino acids, 96% identical to a sequence of 414 amino

CC acids, or 97% identical to a sequence of 242 amino acids, all given in

CC the specification; or a biologically active or immunogenic fragment of

CC the isolated polypeptide. The polypeptides and polynucleotides are useful

CC in diagnosing, treating and preventing diseases or conditions associated

CC with the decreased expression of protein modification and maintenance

CC molecules (PMOD), such as cell proliferative diseases (e.g. cancer,

CC atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,

CC allergies), neurological disorders (e.g. stroke, Parkinson's disease,

CC epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.

CC endometriosis), developmental, vesicle trafficking disorders, and

CC infections (e.g. bacterial, viral, parasitic, protozoal). These are also

CC useful in assessing the effects of exogenous compounds on the expression

CC of nucleic acid and amino acid sequences of PMOD. The PMOD or its

CC fragments are useful in screening compounds for effectiveness as agonist

CC or antagonist of the polypeptides, or in altering the expression of the

CC target polynucleotide and compounds that specifically bind to or modulate

CC the activity of the polypeptide. The microarray is useful in monitoring

CC or measuring protein-protein interactions, drug-target interactions, and

CC gene expression profiles. This polynucleotide sequence represents the DNA

CC encoding a human PMOD protein of the invention

XX

SQ Sequence 2572 BP; 780 A; 489 C; 509 G; 794 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 7,47e-132 Length: 2572

Score: 1240.00 Matches: 239

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

Ds: 8 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x ABR233207 (1-2572)

QY 1 LeuIeuAlaLeuValGInGIyVAlSerAspHeuLeuLeuGlyLysSerLeuAspIle 20

Db 669 CTGTGGCATTAAGTTCACGAAGATTTCACATTTCTGCTTTTGAAAGATCTTTGCACATA 728

QY 21 AlaProGIyASpMetLeuAspLysValAlaArgArgLeuSerLeuIelyHisProGIu 40

Db 729 GCACCAGGTGACATGCTTGGACAAGTGCACAGAAGACTTTTATATTAACATCCAGAG 788

QY 41 CysSerThrMetSerGIyGLyLysAlaIleGIuHisLeuAlalySGInGlyAsnArgPhe 60

Db 789 TGCTTCACCAATGAGTGGTGGGAAAGCCATACAAACATTTGGCCAAACAAAGAAATAGATT 848

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThGIy 80

Db 849 CATTTTGACATCAACCTCCCTCGATCATGCTAATAAATGTGATTTTCTTTATCTGGA 908

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGIuLysGIuGlyIyleGIuLys 100

Db 909 CTTCAACACGCTTACTGATTAATATATATGAAAAAGAAAAGGAAGATTCAGAGAG 968

QY 101 GlyCiniIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisFthMetAlaCys 120

Db 969 GGGCAAAATCTGTCTTCAGACAGACATTCCTGCCACAGTAGACACACAAATGGCARGT 1028

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCylsGlnArgAspLeuLeuPro 140

Db 1029 CATCTTGAAAAGAACACATCGGCGTATTCGTTTTTAAAGCAGAGACTGTGTACTT 1088

QY 141 GlnAsnAsnAlaValAlleuValAlaSerGIyGIyValAlaSerAsnPheTyrlleArgArg 160

Db 1089 CAATTAATATGACGATCTGTGTGCATCTGTGTGTGTGCGAAGTAACTTCTATATCCGCGA 1148

QY 161 AAlaLeuGluIleuThraAsnAlaThrGlnCysThrIleuLeuCySPProProArgLeu 180  
DB 1149 GCTCTGGAATTTTAAACAAACGACAGTGCATTTGTGTGTCTCTCCACAGACTA 1208  
QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluTrpGluLeuArgAlaGlyLeu 200  
DB 1209 TGCACGTATATGCACTATGATGATGCAAGAAAGTATTAAGACATGACGTGCGCTTG 1268  
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyGluProIleCysPProLeuGlyValAsp 220  
DB 1269 GGCATTTCATACATACATGAGAGCATCCGCTATGACCAAAATCTCTTGGAGTAC 1328  
QY 221 IleSerIleGluValGlyGluValAsnIleIleValProGluLeuIleuMetGluIle 239  
DB 1329 ATATCAAAAGAACTTGAGAGAGCTTCATTAAGATACCAATTAATAATGAGAGATA 1385  
RESULT 6  
AAD46856  
ID AAD46856 standard; cDNA; 1820 BP.  
AC AAD46856;  
XX  
XX  
DT 27-JAN-2003 (first entry)  
XX  
XX  
DE Human glycoprotease 28472 cDNA.  
XX  
XX  
KW Human; adenosine deaminase; seven transmembrane domain receptor; cancer;  
7TM; glycoprotease; immune disorder; IGA deficiency; allergy; arrhythmia;  
rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular disorder;  
hypertension; ischemic heart disease; obesity; myocardial infarction;  
endothelial cell disorder; Grave's disease; psoriasis; brain disorder;  
Parkinson's disease; Alzheimer's disease; haematopoietic disorder;  
cerebral oedema; metabolic disorder; liver disorder; platelet disorder;  
chromosome mapping; tissue typing; gene therapy; neuroprotective;  
cystostatic; anorectic; cardiac; haemostatic; gene; ss.  
XX  
XX  
OS Homo sapiens.  
XX  
XX  
FH Key Location/Qualifiers  
FT CDS 146..1390  
FT /tag= a  
FT /product= "Human 28472 protein"  
FT /note= "This region is specifically claimed as SEQ ID NO:  
6 in claim 1 of the specification"  
XX  
XX  
PN WO200274960-A2.  
XX  
XX  
PD 26-SEP-2002.  
XX  
XX  
PF 08-NOV-2001; 2001WO-US051427.  
XX  
XX  
PR 08-NOV-2000; 2000US-0246768P.  
XX  
XX  
PR 08-NOV-2000; 2000US-0246772P.  
XX  
XX  
PR 15-NOV-2000; 2000US-0249185P.  
XX  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
XX  
PI LeiJy KR, Kapeller-Libermann R, Glucksmann M;  
XX  
XX  
DR WPI; 2002-759898/82.  
XX  
XX  
DR P-PSDB; AAE29234.  
XX  
XX  
PT New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,  
PT useful for diagnosing and treating cancer, immune, cardiovascular,  
PT hemetopietic, brain, pain, metabolic, liver or platelet disorders, and  
PT in pharmacogenomics.  
XX  
XX  
PS Claim 1; Fig 8; 178pp; English.  
XX  
XX  
CC The present invention relates to novel 38650, 28472, 5495, 65507, 81588  
CC or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-  
CC protease or seven transmembrane domain (7TM) receptor family members.  
CC Sequences of the invention are useful in diagnosing and treating cancer

CC or aberrant cellular proliferation and/or differentiation (e.g. colon or  
CC lung cancer), immune disorders (e.g. selective IGA deficiency, rheumatoid  
CC arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.,  
CC hypertension, atherosclerosis, arrhythmias, ischemic heart disease,  
CC myocardial infarction, thrombus) including endothelial cell disorders  
CC (e.g. psoriasis, Grave's disease), haematopoietic disorders, brain  
CC disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease),  
CC pain and metabolic disorders (e.g. obesity), liver disorders or platelet  
CC disorders. They are also useful in screening assays, predictive medicine  
CC (e.g. diagnostic assays, prognostic assays, monitoring clinical trials  
CC and pharmacogenetics) and prophylactic and therapeutic methods. The  
CC nucleic acids may also be used in chromosome mapping, tissue typing and  
CC forensic biology and as surrogate markers. Sequences of the invention are  
CC also used in gene therapy. The present sequence is human glycoprotease  
CC 28472 cDNA  
XX  
SQ Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;  
Alignment Scores:  
Pred. No.: 8.22e-128 Length: 1820  
Score: 1203.00 Matches: 232  
Percent Similarity: 98.33% Conservative: 3  
Best Local Similarity: 97.07% Mismatches: 4  
Query Match: 97.02% Indels: 0  
DB: Gaps: 0  
US-10-649-273-2\_COPY\_176\_414 (1-239) x AAD46856 (1-1820)  
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGluGlySerLeuAspIle 20  
DB 671 CTGTGGCATATTAGTTCAAGAGGTTTCAGATTTTCCTTGGAAAGCTTTTGGACATA 730  
QY 21 AlaProGlyAspMetLeuAspIleValAlaArgProLeuSerLeuIleValHisProGlu 40  
DB 731 GCACCAAGTGCATGCTTGAACAGTGCAGAAAGACTTTCTTATTAATCAATCCAGAG 790  
QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaValGlnGlyValAsnArgPhe 60  
DB 791 TGCCTCCACATGACTGGTGGGAAAGCCATGAGACATTTGGCCAAACAGAAATAGATT 850  
QY 61 HisPheAspIleIleYsPProProLeuHisAlaValAsnCyAspPheSerPheThrGly 80  
DB 851 CATTTGACATCAAAACCTCCCTTCATCATGCTAAAAATGTGATTTTCTTTTACTGGA 910  
QY 81 LeuGlnHisValThrAspIleIleIleMetIleYsGluYsGluGluGlyIleGluYs 100  
DB 911 CTTCAACACGTTACTGATTAATAATGATAAACAAGAAACAAGAGAGTATTGAGAG 970  
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120  
DB 971 GGGCAATTCCTGCTTTCAGCAGACATCTCCACAGTACAGACACATGCGCATCT 1030  
QY 121 HisLeuValIysArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuLeuPro 140  
DB 1031 CATCTTGAAAAGAACACATCGGCTATTCTGTTTGTAGACAGAGACTTGTACT 1090  
QY 141 GlnAsnAsnAlaValIleValAlaSerGlyIleValAlaSerAsnPheTyIleArgArg 160  
DB 1091 CAAATATATGCAAGACGTGTGATCGTGTGTGCGCAAGTAACTTCTATATCGCAGA 1150  
QY 161 AAlaLeuGluIleuThraAsnAlaThrGlnCysThrIleuLeuCySPProProArgLeu 180  
DB 1151 GCTCTGGAATTTTAAACAAACGACAGTGCATTTGTGTGTCTCTCCACAGACTA 1210  
QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluTrpGluLeuArgAlaGlyLeu 200  
DB 1211 TGCACGTATATGCACTATGATGATGCAAGAAAGTATTAAGACATGACGTGCGCTTG 1270  
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyGluProIleCysPProLeuGlyValAsp 220  
DB 1271 GGCATTTCATACATACATGAGAGCATCCGCTATGACCAAAATGTCTCTTGGAGTAC 1330  
QY 221 IleSerIleGluValGlyGluValAsnIleIleValProGluLeuIleuMetGluIle 239

DB 1331 ATATCAAAAGAGTGGAGAAAGCTTCATTAAGTACCAATTAATAATGAGATA 1387  
 RESULT 7  
 ACA60887  
 ID ACA60887 standard; cDNA, 1820 BP.  
 AC ACA60887;  
 DT 08-JUL-2003 (first entry)  
 DE Human cDNA 28472 encoding a glycoprotease.  
 KW Human; ss; gene; cancer; aberrant cellular proliferation;  
 KW differentiation; immune disorders; heart disorder; brain disorder;  
 KW cardiovascular disorder; endothelial cell disorder; pain disorder;  
 KW haematopoietic disorder; blood vessel disorder; metabolic disorder;  
 KW liver disorder; platelet disorder; glycoprotease.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 146..1390  
 FT /tag= a  
 FT /product= "Glycoprotease"  
 FT /note= "this CDS is specifically claimed in claim 1"  
 PN US2003009017-A1.  
 PD 09-JAN-2003.  
 PF 08-NOV-2001; 2001US-00012140.  
 XX 08-NOV-2000; 2000US-0246768P.  
 XX 08-NOV-2000; 2000US-0246772P.  
 PR 15-NOV-2000; 2000US-0249185P.  
 PA (LEIB/) LEIBY R R.  
 PA (KAPE/) KAPPELLER-LIBERMANN R.  
 PA (GLUC/) GLUCKSMANN M A.  
 PI Leiby KR, Kapeller-Libermann R, Glucksmann MA;  
 DR MPI; 2003-428888/40.  
 DR P-PDB; AB009569.  
 XX New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid  
 PT molecules, useful for diagnosing, treating cancer, pain, or immune,  
 PT heart, endothelial cell, hematopoietic, blood vessel, brain, metabolic  
 PT and liver disorders.  
 PS Claim 2; Fig 8; 90pp; English.  
 XX The invention relates to an isolated 38650 (encoding adenosine  
 CC deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7  
 CC transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or  
 CC a sequence which is at least 60% identical to the six nucleic acids or  
 CC their open reading frames, fragments of at least 15 nucleotides,  
 CC naturally occurring variants, or a DNA insert of the plasmid deposited  
 CC with the American Type Culture Collection as Accession No. not defined in  
 CC the specification, which encodes the amino acid sequence). Also included  
 CC are a host cell containing the nucleic acids (used to produce the  
 CC proteins), the encoded proteins, an antibody that selectively binds to  
 CC the polypeptide, and identifying a compound that binds to/modulates the  
 CC activity of the polypeptide. The nucleic acid molecules, polypeptides and  
 CC methods are useful for diagnosing, treating cancer, aberrant cellular  
 CC proliferation and/or differentiation, immune disorders, heart disorders,  
 CC cardiovascular disorders including endothelial cell disorders,  
 CC haematopoietic disorders, blood vessel disorders, brain disorders, pain  
 CC and metabolic disorders, liver disorders and platelet disorders (many  
 CC examples of these disorders are given in the specification). The present  
 CC sequence is the Human cDNA 28472 encoding a glycoprotease

SO Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;  
 Alignment Scores:  
 Pred. No.: 8,22e-128 Length: 1820  
 Score: 1203.00 Matches: 232  
 Percent Similarity: 98.33% Conservative: 3  
 Best Local Similarity: 97.07% Mismatches: 4  
 Query Match: 97.02% Indels: 0  
 DB: 8 Gaps: 0  
 US-10-649-273-2\_COPY\_176\_414 (1-239) x ACA60887 (1-1820)  
 QY 1 LeuLeuAlaLeuValGlnGlyValSerAspHeuLeuLeuGlnGlySerLeuAspIle 20  
 Db 671 CTGTTGGCATTAAGTTCAGAGAGGTTTCAGATTTCTGCTTCTGGAAAGCTTTGGACATA 730  
 QY 21 AlaProGlyAspMetLeuAspIleValAlaArgArgLeuSerLeuIleValHisProGlu 40  
 Db 731 GCACCAAGTGCACAGCTTGACAGAGTGCAGAGACCTTTCTTAAATTAACATCCAGG 790  
 QY 41 CysSerThrMetSerGlyGlyValAlaIleGlnHisLeuAlaValGlnGlyAsnArgPhe 60  
 Db 791 TGCCTCACCAAGATGCTGGGAGAAAGCCATAGACATTTGGCCAAACAGAAATGATGATT 850  
 QY 61 HisPheAspIleIleProProLeuHisHisAlaValAsnGlyAspPheSerPheThrGly 80  
 Db 851 CATTTCACATCAAACTCCCTGCATCATCTAAAAATTTGATTTCTTTTACTGCA 910  
 QY 81 LeuGlnHisValIleThrAspIleIleLeuMetLysGlnGlyValGlnGlyIleGlnLys 100  
 Db 911 CTTCAACAGCTTACTGATTAATAATTAAGAAAACAGAAAACAGAGAAAGTATTGAGAG 970  
 QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 120  
 Db 971 GGGCAATTCCTGCTCTTCACGACGACACTTGTCTGCACAGTACAGACACAAATGCGACT 1030  
 QY 121 HisLeuValIleValGlnHisValGlnAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
 Db 1031 CATCTTGAAAAGAACACATCGGGCTATCTGTTTGTAAAGCAGAGACTTGTACT 1090  
 QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaAlaSerAsnPheTyrIleArgArg 160  
 Db 1091 CAAATATATGACAGTACTGCTGTCATCTGTGTGTCGACAGTACTTCATATCCGACGA 1150  
 QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
 Db 1151 GCTTGAAATTTTAACAAACGACACACAGTCACTTGTGTGTCTCTCCAGACTA 1210  
 QY 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
 Db 1211 TGCACCTGATTAATGCGATTATATGATTCATGCAATGATTTGAAGAAGCTACGCTGG 1270  
 QY 201 GlyIleLeuHisAspIleGlnGlyIleArgTyrGlnProLysCysProLeuGlyValAsp 220  
 Db 1271 GGCATTTTACATGACATAGAGGATCCGCTATGAAACCAAAATGTCCTTGGATAGAC 1330  
 QY 221 IleSerLysGluValGlyGlnAlaSerIleLysValProGlnLeuLysMetClnIle 239  
 Db 1331 ATATCAAAAGAGTGGAGAGCTTCATTAAGTACCAATTAATAATGAGATA 1387  
 RESULT 8  
 ABSS7020  
 ID ABSS7020 standard; cDNA, 1821 BP.  
 AC ABSS7020;  
 DT 30-JAN-2003 (first entry)  
 DE cDNA encoding novel human glycoprotease 28472.  
 XX Cancer; aberrant cell proliferation; aberrant cell differentiation;  
 XX breast cancer; ovarian cancer; prostate cancer; colon cancer;  
 KW lung cancer; immune disorder; heart disorder; cardiovascular disorder;

endothelial disorder; hematopoietic disorder; blood vessel disorder;  
 brain disorder; pain; metabolic disorder; liver disorder; diabetes;  
 platelet disorder; carcinoma; sarcoma; leukemia; Hodgkin's disease;  
 autoimmune disorder; hypertension; atherosclerosis; heart failure;  
 myocardial infarction; ischemic heart disease; Crohn's disease;  
 Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;  
 cerebral ischemia; peripheral neuropathy; Alzheimer's disease;  
 Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; gene; ss.  
 Homo sapiens.  
 Key Location/Qualifiers  
 CDS 147..1391  
 /tag= a  
 /product= "glycoprotease 28472"  
 /note= "Specifically claimed in claim 1"  
 WO200272733-A2.  
 03-OCT-2002.  
 08-NOV-2001; 2001WO-US046724.  
 08-NOV-2000; 2000US-0246768P.  
 08-NOV-2000; 2000US-0246772P.  
 15-NOV-2000; 2000US-0249185P.  
 (MILL-) MILLENNIUM PHARM INC.  
 Leiby KR, Kapeller-Libermann R, Glucksmann M;  
 WPI; 2003-029938/02.  
 P-PSDB; ABG71162.  
 New adenosine deaminase, glycoprotease and seven transmembrane domain  
 nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,  
 81588 and 14354, useful for treating e.g. leukemias, Hodgkin's disease or  
 hypertension.  
 Claim 2: Fig 8A-B; 178pp; English.  
 The invention describes isolated 38650, 28472, 5495, 65507, 81588 and  
 14354 nucleic acid molecules (I) and their encoded polypeptides (II). The  
 38650 nucleic acid molecule comprises a sequence encoding adenosine  
 deaminase. The 28472 nucleic acid molecule comprises a sequence encoding  
 a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise  
 sequences that encode a human seven transmembrane domain (7TM). The  
 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide  
 sequences are useful for diagnosing, preventing or treating a subject  
 with or at risk of developing a disorder, e.g. cancer or aberrant  
 cellular proliferation and/or differentiation (e.g. breast, ovarian,  
 prostate, colon or lung cancer), immune disorders, heart disorders,  
 cardiovascular disorders, endothelial disorders, hematopoietic disorders,  
 blood vessel disorders, brain disorders, pain and metabolic disorders,  
 liver disorders or platelet disorders. These disorders include carcinoma,  
 sarcoma, leukemia, Hodgkin's disease, autoimmune disorders,  
 hypertension, atherosclerosis, heart failure, myocardial infarction,  
 ischemic heart disease, Crohn's disease, Grave's disease, Kawasaki  
 syndrome, Raynaud's disease, aneurysm, cerebral ischemia, peripheral  
 neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,  
 cachexia or diabetes. This sequence encodes the novel human glycoprotease  
 28472  
 XX  
 CC Sequence 1821 BP; 543 A; 365 C; 394 G; 518 T; 0 U; 1 Other;  
 Alignment Scores:  
 Pred. No.: 8, 22e-128 Length: 1821  
 Score: 1203.00 Matches: 232  
 Percent Similarity: 98.33% Conservative: 3  
 Best Local Similarity: 97.07% Mismatches: 4  
 Query Match: 97.02% Indels: 0  
 DB: 10 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x ABS57020 (1-1821)  
 QY 1 LeuLeuAlaLeuValGlnGlyValSerAspSphLeuLeuGlyLysSerLeuAspIle 20  
 Db 672 CTGTGGCATTAGTTCAGAGAGGTTTCAGATTTCCTGCTTGGAAAGCTTTGGACATA 731  
 QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
 Db 732 GCACCAAGTGCACAGCTTGCACCAAGTGCACCAAGTGCACCAAGTGCACCAAGTGCAC 791  
 QY 41 CysSerThrMetSerGlyValLysAlaIleGlnHisLeuAlaLysGlnGlyValAsnArgPhe 60  
 Db 792 TGCTCCACAGTGTGGTGGAAAGCCATTGAAACATTGGCCAAACAGAAATGATTT 851  
 QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
 Db 852 CATTTCACATCAAACTCCCTTCATCAGCTTAAATTTGATTTTCTTTACTGCA 911  
 QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlyLysGlyLysIleGlnLys 100  
 Db 912 CTTCAACACGTTACTGATTAATAATGAAACAGAAACAGAAAGGATTTAGAGAG 971  
 QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaLys 120  
 Db 972 GGGCAAACTCTGCTTTCAGCAGCAGATTCCTCCACAGTACAGCACAATGCGCATGT 1031  
 QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140  
 Db 1032 CATCTTGTGAAAAGAACATCGGCTATCTCTGTTTGTAAAGCAGAGACTTGTACT 1091  
 QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaLysSerAsnPheTyrIleArgArg 160  
 Db 1092 CAATATATGACAGTACGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1151  
 QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180  
 Db 1152 GCTCTGGAATTTTAAACAAAGCAGACAGTCTGTGTGTCTCTCTCCACAGACTA 1211  
 QY 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
 Db 1212 TGCACTATATGACATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 1271  
 QY 201 GlyIleLeuHisAspIleGlyLysIleArgTyrGluProLysCysProLeuGlyValAsp 220  
 Db 1272 GGCATTTTACATGACATGAAAGGATCGCATGTAACCAAAATGCTCTTGGAGTAGAC 1331  
 QY 221 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239  
 Db 1332 ATATCAAAAGAACTTGGAGAAAGCTTCCATTAAGTACCAATTAATAATGAGAGATA 1388  
 RESULT 9  
 ADA52832  
 ID ADA52832 standard; cDNA; 2208 BP.  
 XX  
 AC ADA52832;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Human coding sequence, SEQ ID 400.  
 XX  
 KW Cytostatic; Anti-inflammatory; Osteoprotic; Neuroprotective; Nootropic;  
 KW Gene Therapy; human; secretory protein; membrane proteins; cancer;  
 KW inflammatory disease; osteoporosis; neurological disease; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN BP1293569-A2.  
 XX  
 PD 19-MAR-2003.  
 XX  
 PF 21-MAR-2002; 2002EP-00006586.  
 XX  
 PR 14-SEP-2001; 2001JP-00328381.

24-JAN-2002; 2002US-0350435P.  
 PR XX (HELI-) HELIX RES INST.  
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
 XX  
 PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishi S;  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
 PI Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuno Y;  
 DR WPI: 2003-395539/38.  
 DR P-PSDB; ADA54471.  
 XX  
 PT New polynucleotides encoding full-length polypeptides, e.g. secretory  
 PT and/or membrane proteins, useful for developing medicines for diseases in  
 PT which the gene is involved, or as target molecules for gene therapy.  
 XX  
 PS Claim 1, SEQ ID NO 400; 205pp; English.  
 XX  
 CC The present invention relates to novel human secretory or membrane  
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
 CC ADA54071). The coding sequences are useful in the gene therapy of  
 CC diseases caused by abnormalities of the proteins, e.g. cancer,  
 CC inflammatory diseases, osteoporosis or neurological disease.  
 XX  
 SQ Sequence 2208 BP; 660 A; 454 C; 451 G; 643 T; 0 U; 0 Other;  
 XX  
 Alignment Scores:  
 Pred. No.: 3,88e-111 Length: 2208  
 Score: 1059.00 Matches: 211  
 Percent Similarity: 88.28% Conservative: 0  
 Best Local Similarity: 88.28% Mismatches: 4  
 Query Match: 85.40% Indels: 24  
 DB: 10 Gaps: 1  
 US-10-649-273-2\_COPY\_176\_414 (1-239) x ADA52832 (1-2208)  
 QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlySerLeuAsp11e 20  
 DB 869 CTGTTGGCATTACTTCAAGAGTTTCAGATTTCCTCTTGGAAAGCTTTTGGACATA 928  
 QY 21 AlaProGlyAspMetLeuAspLeuValAlaArgArgLeuSerLeuIleYsHsProGlu 40  
 DB 929 GCACCGAGGTGACATGCTTGACAGAGTGGCAAGAACTTCCTTAATTAACATCCAGAG 988  
 QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaValGlnGlyAsnArgPhe 60  
 DB 989 TGCCTCCACCATGAGTGGTGGAAAGCCATTAAGCATTTGGCAAAACAAGAAATAGATT 1048  
 QY 61 HisPheAspIleLeuProPheLeuHisAlaValAsnCysAspPheSerPheThrGly 80  
 DB 1049 CATTTCGACATCAAACTCCCTTCGATCATGCTAAATAATGTGATTTCTTTACTGGA 1108  
 QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGly11eGluLys 100  
 DB 1109 CTTCAACAGCTTACTGATTAATAATTAATTAAGAAAAGAGAAAGATTTGAGAGAG 1168  
 QY 101 GlnGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120  
 DB 1169 GGGCAATCTCTGCTTACGACAGACATTCCTCCACAGTACGACACACAATGCGCATGT 1228  
 QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnAlaGspLeuLeuPro 140  
 DB 1229 CATCTGTGAAAAGAACACATCGGGCTATCTCTTTGTTAAGCAGAGACTCTTACT 1288  
 QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValValAlaSerAsnPheYrIleArgArg 160  
 DB 1289 CAATAAATATGACAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1348  
 QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProPheArgLeu 180  
 DB 1349 GCTCTGGAATTTTAAACAAACGACAGTGCATTTGTTGTCTCTCTCCACGACTA 1408  
 QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200

DB 1409 TGCACATATATGACATTATATGCA----- 1435  
 QY 201 GlyIleLeuHisAspIleGluGlyIleArgYrGluProLysCysProLeuGlyValAsp 220  
 DB 1436 -----TATGTCTCTTGGAGTAAAC 1456  
 QY 221 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239  
 DB 1457 ATATCAAAAGAGTGGAGAAAGCTTCATTAAGTACACACATTAATAATGAGATA 1513  
 RESULT 10  
 ADQ24627  
 ID ADQ24627 standard; DNA; 2890 BP.  
 XX  
 AC ADQ24627;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 7447.  
 XX  
 KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
 KW ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2004048938-A2.  
 XX  
 PD 10-JUN-2004.  
 XX  
 PF 26-NOV-2003; 2003WO-US038193.  
 XX  
 PR 26-NOV-2002; 2002US-0429739P.  
 XX  
 PA (PROT-) PROTEIN DESIGN LABS INC.  
 PI Aziz N, Ginsburg WM, Zlotnick A;  
 DR WPI: 2004-441208/41.  
 XX  
 PT Early detection of soft tissue sarcoma comprises determining expression  
 PT of a gene in a first soft tissue sample and a normal soft tissue sample  
 PT and comparing the gene expression, also useful in treating soft tissue  
 PT sarcoma.  
 XX  
 XX Example 2; SEQ ID NO 7447; 210bp; English.  
 PS  
 CC The invention relates to a novel method for detecting soft tissue sarcoma  
 CC which comprises obtaining a first soft tissue sample from an individual  
 CC and a normal soft tissue sample from the same or different individual,  
 CC determining the expression of a gene in both samples and comparing the  
 CC expression of the gene in both soft tissue samples, where a higher level  
 CC of protein expression in the first soft tissue sample indicates the  
 CC presence of soft tissue sarcoma. The method of the invention has  
 CC cyrostatic applications and may be useful for detecting soft tissue  
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
 CC acid sequences may be useful in diagnostic and screening applications.  
 CC The current sequence is that of a human soft tissue sarcoma-upregulated  
 CC DNA of the invention. The current sequence is not shown within the  
 CC specification per se but was submitted in CD format by the inventor.  
 XX  
 SQ Sequence 2890 BP; 869 A; 609 C; 611 G; 789 T; 0 U; 12 Other;  
 XX  
 Alignment Scores:  
 Pred. No.: 5.67e-111 Length: 2890  
 Score: 1059.00 Matches: 211  
 Percent Similarity: 88.28% Conservative: 0  
 Best Local Similarity: 88.28% Mismatches: 4  
 Query Match: 95.40% Indels: 24  
 DB: 12 Gaps: 1  
 US-10-649-273-2\_COPY\_176\_414 (1-239) x ADQ24627 (1-2890)

QY 1 LeuLeuAlaLeuValGlnGlyValIleSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
 DB 1526 CCGTGGCACTGATGTTAAGAGATTTTCAGATTTCTGCTTGGAAAGCTTTGGACATA 1585  
 QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysSHIAspGlu 40  
 DB 1586 GCAACGAGTGCATGCTTGAACAGGTGGCAAGAGACTTCTTTATATTAACATCCAGAG 1645  
 QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60  
 DB 1646 TCGTCACCATGATGCTGTGGAAAGCATTAAGACATTTGGCCAAAGAAAGAAATAGATT 1705  
 QY 61 HisPheAspIleLysProPheLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80  
 DB 1706 CATTTTGACATCAAACTCCCTTCATCATGCTAAATAATGATGATTTCTTTTACTGGA 1765  
 QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIleGluLys 100  
 DB 1766 CTTCAACAGCTTATCTATTAATAATTAATGAAGAAAGAAAGAAAGAAATGAGAG 1825  
 QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120  
 DB 1826 GGGCAAAATCTGCTCTCAGACAGACATTCCTGCCACATACAGCAACAAATGGCATGT 1885  
 QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140  
 DB 1886 CATCTTGAAAAAGAACACATCGGGCATTCCTGTTTGTAGCAGAGACTTGTACTCT 1945  
 QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrlleArgArg 160  
 DB 1946 CAAATAAATGCACTGCTGTGATCTGGGCTGCCAAGTAATCTGTATCCGAGAGA 2005  
 QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
 DB 2006 GCTCTGGAATTTTAAACAAGCAACACAGTCTGTTGTGCTCTCCAGCAGCTA 2065  
 QY 181 CysTrpAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
 DB 2066 TCCACTGATTAATGGCATTTATGATTGCA----- 2092  
 QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrlleGluProLysCysProLeuGlyValAsp 220  
 DB 2093 -----TGATGTCCTCTTGGAGTGAAC 2113  
 QY 221 IleSerLysGluValGlyGlnAlaSerIleLysValAlaProGlnLeuLysMetGlnIle 239  
 DB 2114 ATATCAAAAGAAAGTTGAGAGAGCTTCATTAAGTACCAATTAATAAATGAGATA 2170  
 RESULT 11  
 ADE31345/c  
 ID ADE31345 standard; DNA; 3358 BP.  
 AC ADE31345;  
 XX  
 DE 29-JAN-2004 (first entry)  
 XX  
 XX Human diagnostic and therapeutic polynucleotide (dithp), SEQ ID NO 100.  
 XX  
 XX diagnostic and therapeutic polynucleotide; dithp; antiarteriosclerotic;  
 XX antiinflammatory; cerebroprotective; antilipemic; antidiabetic;  
 XX immunosuppressive; neuroprotective; nootropic; neuroleptic; tranquilizer;  
 XX osteopathic; antirheumatic; antirheumatic; cyrostatic; hepatotropic;  
 XX virucide; haemostatic; anti-HIV; antithyroid; thyromimetic;  
 XX dermatological; antibacterial; fungicide; antiparasitic; anticonvulsant;  
 XX thrombolytic; anticoagulant; anorectic; vasotrophic; antitumor;  
 XX gene therapy; protein replacement therapy; human; gene; ds.  
 OS Homo sapiens.  
 XX  
 XX WO2003062376-A2.  
 XX  
 PD 31-JUL-2003.

PF 13-JAN-2003; 2003WO-US001096.  
 XX  
 PR 16-JAN-2002; 2002US-0349384P.  
 PR 17-JAN-2002; 2002US-0349413P.  
 PR 17-JAN-2002; 2002US-0349946P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Jones AL, Dahl CR, Gietzen D, Chinn J, Dufour GB, Jackson JL,  
 PI Yu JY, Tuason O, Yap PE, Amshy SR, Dam TC, Liu TF, Gerstin EH,  
 PI Peralta CH, Lewis SA, Chen A, Marwaha R, Lan KY, Urashka ME,  
 PI Kristam SR, Kolluru V, Panesar IS;  
 XX  
 DR WPI; 2003-636732/60.  
 DR P-PSDB; ADE31156.  
 XX  
 PT New human diagnostic and therapeutic polynucleotides and polypeptides,  
 PT useful for diagnosing, treating or preventing e.g. leukemia, brain  
 PT cancer, arteriosclerosis, AIDS, thyroiditis, infections, obesity, stroke  
 PT or Alzheimer's.  
 PS Claim 1; SEQ ID NO 100; 634pp; English.  
 XX  
 CC The invention relates to a novel isolated human diagnostic and  
 CC therapeutic polynucleotide (designated dithp). The novel dithp  
 CC polynucleotide comprises: any of 188 DNA sequences consisting of 195-7798  
 CC base pairs fully defined in the specification; a polynucleotide  
 CC comprising a naturally occurring polynucleotide sequence at least 90%  
 CC identical to the dithp polynucleotide; a polynucleotide complementary to  
 CC the dithp polynucleotide or its polynucleotide which is at least 90%  
 CC identical; or an RNA equivalent of any of the polynucleotides mentioned  
 CC above. The dithp polynucleotides have the following activities:  
 CC antiarteriosclerotic, antiinflammatory, cerebroprotective, antilipemic,  
 CC antidiabetic, immunosuppressive, neuroprotective, nootropic, neuroleptic,  
 CC tranquilizer, osteopathic, antirheumatic, antirheumatic, cyrostatic,  
 CC hepatotropic, virucide, haemostatic, anti-HIV, antithyroid, thyromimetic,  
 CC dermatological, antibacterial, fungicide, antiparasitic, anticonvulsant,  
 CC thrombolytic, anticoagulant, anorectic, vasotrophic, and antitumor. The  
 CC novel dithp polynucleotides polypeptide can be used in gene therapy and  
 CC protein replacement therapy. The dithp polynucleotides or dithp  
 CC polypeptides are useful for diagnosing, preventing or treating diseases  
 CC associated with the expression of human molecules. In particular, these  
 CC diseases include cancers (e.g. adenocarcinoma, leukaemia, melanoma, brain  
 CC cancer, breast cancer, cervix cancer, bone cancer, liver cancer, lung  
 CC cancer) or other cell proliferative disorders (e.g. arteriosclerosis,  
 CC atherosclerosis, cirrhosis, hepatitis, polycythemia vera, primary  
 CC thrombocytopenia), autoimmune/inflammatory disorders (e.g. AIDS,  
 CC Addison's disease, thyroiditis, Crohn's disease, Graves' disease,  
 CC Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid  
 CC arthritis, scleroderma, systemic lupus erythematosus), infections (e.g.  
 CC viral, bacterial, fungal or parasitic infection), developmental disorders  
 CC (e.g. Cushing's syndrome or epilepsy), endocrine disorders (e.g.  
 CC thrombosis, hypopituitarism, hypogonadism, gigantism, goiter) metabolic  
 CC disorders (e.g. hypercholesterolaemia, hypoglycaemia, diabetes,  
 CC hyperlipidaemia, obesity), neurological disorders (e.g. ischaemic  
 CC cerebrovascular disease, stroke, Alzheimer's disease, Pick's disease,  
 CC Huntington's disease, Parkinson's disease, Creutzfeldt-Jakob disease,  
 CC anxiety, schizophrenia), gastrointestinal disorders (e.g. ulcers),  
 CC transport disorders (e.g. akinesia or multidrug resistance), or  
 CC connective tissue disorders (e.g. Paget's disease or rickets). This  
 CC polynucleotide sequence represents one of the human dithp DNA sequences  
 CC of the invention.  
 XX  
 SQ Sequence 3358 BP; 1105 A; 577 C; 601 G; 1075 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 1,4e-103 Length: 3358  
 Score: 995.50 Matches: 209  
 Percent Similarity: 87.82% Conservative: 0  
 Best Local Similarity: 87.82% Mismatches: 2  
 Query Match: 80.28% Indels: 29  
 DB: 10 Gaps: 1

US-10-649-273-2\_copy\_176\_414 (1-239) x ADE31345 (1-3358)

```

Oy 29 ValAlaArgArgLeuSerLeuIleLeuHisProGluCysSerThrMetSerGly-GlyLeu 48
Db 2104 GTGGCAGAGAGACTTTCTTTTAAATCAATCCAGAGTCTCCACATGACGTGGGAA 2045
Oy 48 salaiIegluHisleuAlaIysGlnGlyAsnArgPheHisPheAspIleLeuProPhe 68
Db 2044 AGCCATGAGACATTGGCCAAACAGAGAAATAGATTTCATTTGACATCAAACTCCCTT 1985
Oy 68 uHisHisAlaIysAlaCysAspPheSerPheThrGlyLeuGlnHisValThrAspIle 88
Db 1984 GCATCAACGCTAAATAATGTGATTTTCTTTTACGACTTCAACACGTTACTGATAAAT 1925
Oy 88 eIleMetIleIysGluIleIysGluIle----- 96
Db 1924 AATTAATGAAAAGAGAAAAGAGAAAGTATATTCTAATTAGTAAAGTTGAACAGATATA 1865
Oy 97 -----GlyIleGluIysGluIle 103
Db 1864 TATTCTGCAATGTGCTTAAATAATAGCTCATTTCTGACGATTTGAGAAAGGGGAAA 1805
Oy 103 IeLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisIleuV 123
Db 1804 TCCGTCTTGCAGCAGACAGACATTGCTGCGCAGATACAGACACAAATGCGCATGCTTG 1745
Oy 123 alIysArgThrHisArgAlaIleLeuPheCysIleGlnArgAspLeuIleuProGlnAsnA 143
Db 1744 TGAATAACACACATCGGCTATTTCTGTTTGTAAAGAGAGACTGTGTACCTCAAAATA 1685
Oy 143 snAlaValleuValAlaSerGlyIleValAlaSerAsnPheThrIleArgArgAlaLeuG 163
Db 1684 ATGCAGTACTGGTTCATCTGT-GGTGCGCAAGTAACTTTATATCCGACAGCTCTCG 1626
Oy 163 IuIleLeuThrAsnAlaThrGlnCysThrLeuIleuCysProProAlaArgLeuCysThrA 183
Db 1625 AATTTTAAAC-AACGCAACACAGTGCCTTTGTTGTGTCCTCCGCCAGACTATGACATG 1567
Oy 183 sPhaGngIleIleMetIleAlaTrpAsnGlyIleGluIysGluLeuArgAlaGlyLeuGlyIle 203
Db 1566 ATTAATGCATATATGATTCATGAGATGTATTTGAAGACTGCTGCTGGGCATATT 1507
Oy 203 euHisAspIleGluGlyIleArgGlyIleArgProIysCysProLeuGlyValAspIleSer 223
Db 1506 TACATGACATAGAGGACCTGCTATGAAACCAAAATGCTCTTGGAGTACGACATATCA 1447
Oy 223 ysgIuValIglYglAlaSerIleIleIysValProGlnLeuIleMetGluIle 239
Db 1446 AAGAAAGTTGGAGAGCTTCCATTAAGTACCAATTAATAATGAGATTA 1397

```

RESULT 12  
ABQ75508  
ID ABQ75508 standard; DNA; 1572 BP.

AC ABQ75508;  
XX  
XX 07-NOV-2002 (first entry)  
XX  
XX Murine sialoglycoprotease-like gene sequence SEQ ID NO:7.  
DE  
XX  
XX Murine, mouse; protease; calcium activated neutral protease type 5;  
KM CAPN5; trypsin 4; sialoglycoprotease; enzyme; genetic disease;  
KM neurological; neuropsychological; psychotic illness; transgenic animal;  
KM gene; ds.  
XX  
XX Mus musculus.  
XX  
XX PN W0200245491-A2.  
XX  
XX PD 13-JUN-2002.  
XX  
XX PF 05-DEC-2001; 2001WO-US046405.

PR 06-DEC-2000; 2000US-0251803P.  
PR 06-DEC-2000; 2000US-0251820P.  
PR 13-DEC-2000; 2000US-0255971P.  
PA (DELTA-) DELTAGEN INC.  
PI Allen KD, Leviten MM;  
DR WPI; 2002-657389/70.  
XX  
XX Novel transgenic animal, comprising a disruption in protease target gene,  
PT is useful for identifying agents that ameliorates a phenotype associated  
PT with a disruption in a protease target gene.  
XX  
XX Example 3; Fig 7; 62pp; English.  
XX  
XX The present invention describes a non-human transgenic animal (I)  
CC comprising a disruption in a protease target gene (PG) selected from  
CC calcium activated neutral protease type 5 (CAPN5) gene, trypsin 4 gene  
CC and sialoglycoprotease-like gene. Also described is a targeting construct  
CC (II), comprising a first polynucleotide sequence homologous to at least a  
CC first portion of PG, a second polynucleotide sequence homologous to at  
CC least a second portion of PG and a selectable marker. (II) is useful for  
CC producing a transgenic mouse comprising a disruption in a protease target  
CC gene, by introducing (II) into a cell, introducing the cell into a  
CC blastocyst, implanting the resulting blastocyst into a pseudopregnant  
CC mouse, where the pseudopregnant mouse gives birth to a chimeric mouse,  
CC and breeding the chimeric mouse to produce the transgenic mouse. (I) is  
CC useful for identifying an agent that modulates the expression or function  
CC of a protease target gene, by administering an agent to (I) and  
CC determining whether the expression or function of the disrupted protease  
CC target gene in (I) is modulated. (I) is also useful for testing the  
CC efficacy of proposed genetic and pharmacological therapies for human  
CC genetic diseases, such as neurological, neuropsychological or psychotic  
CC illness. The present sequence represents murine sialoglycoprotease-like  
CC gene sequence, which is used in an example from the present invention  
SQ

Sequence 1572 BP; 459 A; 337 C; 340 G; 429 T; 0 U; 7 Other;

Alignment Scores:  
Pred. No.: 1.28e-89 Length: 1572  
Score: 870.00 Matches: 184  
Percent Similarity: 82.77% Conservative: 13  
Best Local Similarity: 77.31% Mismatches: 37  
Query Match: 70.16% Indels: 5  
DB: Gaps: 2

US-10-649-273-2\_COPY\_176\_414 (1-239) x ABQ75508 (1-1572)

```

Oy 1 LeuLeuAlaLeuValGluGlyValaSerAspPheLeuLeuGlyIysSerLeuAspIle 20
Db 528 CTGTTGAKRTTAAAGTCCAGAGTTCCTGCTCTTGGAGAGCTTTGGACATA 587
Oy 21 AlaProGlyAspMetLeuAspIysValAlaArgArgLeuSerLeuIleHisProGlu 40
Db 588 GCACAGGCGACATGCTTGCACAGTGGCAAAAGACTTTCTTAATCAATCCAGAA 647
Oy 41 CysSerThrMetSerGlyIleIysAlaIleGluHisIleuAlaIysGlnIysAsnArgPhe 60
Db 648 TGTTCACATAGAGTGTGGAAGAGCTATAGAAAGTGTGGCAAGACGGAATAGATTTC 707
Oy 61 HisPheAspIleIysProProLeuHisHisAlaIysAsnArgCysAspPheSerPheThrGly 80
Db 708 CATTTTACTATCAATCCACTATGCAAGATGTCAAGAAATGTGCAATTTTCTTACCGGGA 767
Oy 81 LeuGlnHisValThrAspIleIleIleMetIleIysGluIleGluGlyIleGluIys 100
Db 768 CTTCAACATATTACTGATAGCTAATACACACAGAGAAAAGAGAGCATTTGAGAG 827
Oy 101 GlyIleIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120
Db 828 GGGCAATCTGTCTATCAGCTGACAGACATTTGCTGCGGTACAGATCAACAGGTGTC 887

```







|    |   |
|----|---|
| KM | gene potential; multi-lineage; cell commitment; haematopoietic stem cell; |
| KW | HSC; multipotent progenitor; MP; common lymphoid progenitor; CLP;         |
| KX | common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; ds  |
| OS | Mus sp.   |
| XX | WO2003093445-A2.  |
| PN | 13-NOV-2003.  |
| PD | 05-MAY-2003; 2003WO-US014114.   |
| PF | 03-MAY-2002; 2002US-0377383P.   |
| XX | (STOW-) STOWERS INST MEDICAL RES.   |
| PA | L1 L;   |
| XX | WPI; 2004-022656/02.  |
| DR |   |
| XX |   |
| PT | Classifying an unknown multi-lineage affiliated gene comprises isolating  |
| PT | expressed nucleic acid sequences from the discrete cell sub-populations.  |
| PS | Claim 8; SEQ ID NO 3118; 123pp; English.                                  |

This invention relates to a novel method for predicting gene potential by associating nucleic acid sequences of unknown function with particular sub-population profiles. Specifically, it refers to classifying an unknown multi-lineage affiliated gene by collecting hybridisation data to develop a gene expression map, in order to determine the discrete sub-population where it is expressed. The present invention describes methods for predicting the lineage commitment of genes associated with the self-renewing haematopoietic (blood) stem cells (HSCs), as well as the non-self-renewing multipotent progenitors (MPPs), common lymphoid progenitors (CLPs) and common myeloid progenitors (CMPs), which are collectively referred to as bone marrow stem cells populations. As such, these methods can be used to identify associated multi-lineage affiliated genes and hence the underlying molecular mechanisms in physiological haematopoietic development. This polynucleotide sequence is DNA associated with a murine MPP sub-population of cells of the invention.

**SQ** Sequence 371 BP; 104 A; 75 C; 84 G; 92 T; 0 U; 16 Other;

|                        |          |
|------------------------|----------|
| Alignment Scores:      |          |
| Pred. No.:             | 2,75e-44 |
| Score:                 | 468.00   |
| Percent Similarity:    | 78.86%   |
| Best Local Similarity: | 74.80%   |
| Query Match:           | 37.74%   |
| DB:                    | 12       |
|                        |          |
| Length:                | 377      |
| Matches:               | 92       |
| Conservative:          | 5        |
| Mismatches:            | 26       |
| Indels:                | 0        |
| Gaps:                  | 0        |

US-10-649-273-2\_COPY\_176\_414 (1-239) X ADL86725 (1-371)

|    |     |  |     |
|----|-----|--|-----|
| Qy | 109 | asplleialaalatThValGlnhiStrHmecaIaCySHiSleuValIySaAGTThriSaG | 128 |
| Db | 2   | GACATTCCTGCTCGCGGTACGACATGCACAGCGTGCACCTTCGCGAAGAACACATCGC   | 61  |
| Qy | 129 | AlaIleleuPheCysIySGlnArgAspleuLeuProGlnaSnanaIaValIleuValaIa | 148 |
| Db | 62  | GCTATTCTGTTTGGACAGACAGAAATTTGCTCTCCAGCTAACGACAGATTATGTTGTA   | 121 |
| Qy | 149 | SerGIyGIyValAlaSeranPheYrIIeaArgAlaIleuGIyIleuThraPnaIa      | 166 |
| Db | 122 | TCTGAGGTGTTGCAGTAACTTGTACATCCGAAAGCATGGAAATTGTCGCAATGCA      | 181 |
| Qy | 169 | ThrgInCysThrIleuLeuCysProProArgLeuCyStrAspAngIyIIeMetIle     | 188 |
| Db | 182 | ACGCAGTGCACCTTTGTTGTGCCACTCCAAAGCTGTGCATCAACAATGCATCATGATT   | 241 |
| Qy | 189 | AlaITPaanGIyIIegIuArgLeuArgAlaGIyLeuGIyIleuHiSaapIIeGluGIy   | 208 |
| Db | 242 | GCATGAATGGAAATGAAAGATTACGTGCAGNCTTGGCACTTTTACCTGAAANTAGAAAC  | 301 |

|    |  |     |
|----|--|-----|
| Oy | IleatgTGTGCTProlLysCysProLeugIYValAspILeSerLYscgluaValGLylGIuaLa | 228 |
| Dc |  |     |
| Oy | ATCCNNATTGNACCAAAATNTCNTTTTGAGTAGACAATTCAGAGAAGCTCGAAGCT         | 361 |
| Dc | ::  :  |     |
| Oy | SerILEtLyS  231  |     |
| Dc | :::  |     |
| Oy | GccATPADA  370   |     |
| Dc | :  |     |

RESULT 15  
ADL86726  
ID ADL86726 standard; DNA; 371 BP.

**DNA up-regulated in murine multipotent progenitor cells SeqID 3119.**

KW gene potential; multi-lineage; cell commitment; haematopoietic stem cell;  
 KW HSC, multipotent progenitor; MPP; common lymphoid progenitor; CLP;  
 KW common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; ds

|    |                  |
|----|------------------|
| OS | Mus sp.          |
| XX |                  |
| PN | W02003093445-A2. |

PD 13-NOV-2003.  
XX  
PF 05-MAY-2003; 2003MO-US014114.  
XX  
XX 03-MAY-2002; 2002US-0377383P.  
PR  
XX  
PA (STOW-) STOWERS INST MEDICAL RES

PL L1 L;  
XX  
XX WPT; 2004-022656/02.  
DX  
XX  
XX Classifying an unknown multi-lineage affiliated gene comprises isolating  
PT expressed nucleic acid sequences from the discrete cell sub-populations.  
XX  
XX  
PS Claim 8; SEQ ID NO 3119; 123bp; English.

**Sequence** 371 BP; 104 A; 75 C; 84 G; 92 T; 0 U; 16 Other;

|                        |          |
|------------------------|----------|
| Alignment Score:       |          |
| Pred. No.:             | 2.75e-44 |
| Score:                 | 468.00   |
| Percent Similarity:    | 78.86%   |
| Best Local Similarity: | 74.80%   |
| Query Match:           | 37.74%   |
| DB:                    | 12       |
| Length:                | 371      |
| Matches:               | 92       |
| Conservative:          | 5        |
| Mismatches:            | 26       |
| Indels:                | 0        |
| Gaps:                  | 0        |

US-10-649-273-2\_COPY\_176\_414 (1-239) x ADL86726 (1-371)

Qy 109 AspielaalThrValGlnIeThrMetAlaCysHisIeuValLysArgThnHisArg 128  
|||||



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Comphen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: June 16, 2005, 18:30:49 ; Search time 118.721 Seconds

(without alignments)  
3294.036 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414

Perfect score: 1240  
Sequence: 1 ILIATVGVSDFFLLGKSLDI.....DISKEVGEASIKVPQKMEI 239

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1202784 segs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seg length: 0

Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+.p2n.model -DEV=xlp  
-Q=/cgnt2\_1/USPTO.spool\_p/US10649273/runat\_15062005\_111417\_6057/app.query.fasta\_1.1429  
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rnt -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.ccd1  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=p2n -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10649273\_QCGN1\_1\_177\_@runat\_15062005\_111417\_6057 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA.\*  
1: /cgnt2\_6/prodata/1/ina/5A.COMB.seg.\*  
2: /cgnt2\_6/prodata/1/ina/5B.COMB.seg.\*  
3: /cgnt2\_6/prodata/1/ina/6A.COMB.seg.\*  
4: /cgnt2\_6/prodata/1/ina/6B.COMB.seg.\*  
5: /cgnt2\_6/prodata/1/ina/PCTUS.COMB.seg.\*  
6: /cgnt2\_6/prodata/1/ina/backfile1.seg.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length  | DB ID | Description                          |
|------------|--------|-------------|---------|-------|--------------------------------------|
| 1          | 1240   | 100.0       | 1416    | 4     | US-09-774-528-177 Sequence 177, App  |
| 2          | 1240   | 100.0       | 1526    | 4     | US-10-067-443-23 Sequence 23, Appl   |
| 3          | 1240   | 100.0       | 2197    | 4     | US-10-067-443-1 Sequence 21, Appl    |
| 4          | 1213.5 | 97.9        | 1387    | 4     | US-10-067-443-21 Sequence 21, Appl   |
| 5          | 1213.5 | 97.9        | 1387    | 4     | US-10-067-443-20 Sequence 20, Appl   |
| 6          | 248.5  | 58.5        | 14364   | 4     | US-09-596-002-38 Sequence 38, Appl   |
| 7          | 247    | 19.3        | 1053    | 4     | US-09-540-236-806 Sequence 806, App  |
| 8          | 231    | 18.6        | 1059    | 4     | US-09-252-991A-884 Sequence 884, App |
| 9          | 231    | 18.6        | 1206    | 4     | US-09-252-991A-801 Sequence 801, App |
| 10         | 224    | 18.1        | 1830121 | 4     | US-09-557-884-1 Sequence 1, Appl     |
| 11         | 224    | 18.1        | 1830121 | 4     | US-09-543-990A-1 Sequence 1, Appl    |
| 12         | 221.5  | 17.9        | 996     | 4     | US-09-902-540-6612 Sequence 6612, Ap |

|   |    |       |      |         |   |                     |
|---|----|-------|------|---------|---|---------------------|
| C | 13 | 221.5 | 17.9 | 2582    | 4 | US-09-902-540-503   |
|   | 14 | 212   | 17.1 | 1074    | 4 | US-09-543-661A-2341 |
|   | 15 | 205   | 16.5 | 1032    | 4 | US-09-489-039A-2050 |
|   | 16 | 200   | 16.1 | 1315    | 1 | US-08-087-797-1     |
|   | 17 | 197   | 15.9 | 1092    | 4 | US-09-107-532A-2955 |
|   | 18 | 190   | 15.3 | 1008    | 3 | US-08-987-121A-5    |
|   | 19 | 190   | 15.3 | 1011    | 3 | US-08-987-121A-3    |
|   | 20 | 189   | 15.2 | 822     | 4 | US-09-710-279-127   |
|   | 21 | 189   | 15.2 | 3993    | 4 | US-09-710-279-3985  |
|   | 22 | 187   | 15.1 | 1155    | 4 | US-09-602-777A-147  |
|   | 23 | 186.5 | 15.0 | 4403765 | 3 | US-09-103-840A-2    |
|   | 24 | 186.5 | 15.0 | 4411529 | 3 | US-09-103-840A-1    |
|   | 25 | 186   | 15.0 | 1006    | 3 | US-08-961-083-51    |
|   | 26 | 186   | 15.0 | 1006    | 4 | US-09-536-784-51    |
|   | 27 | 186   | 15.0 | 1107    | 3 | US-09-134-001C-1072 |
|   | 28 | 186   | 15.0 | 10974   | 3 | US-08-961-527-214   |
| C | 29 | 184   | 14.8 | 1011    | 3 | US-09-066-512-1     |
|   | 30 | 182   | 14.7 | 1011    | 4 | US-09-583-110-2196  |
|   | 31 | 182   | 14.7 | 1663    | 4 | US-09-620-312D-6    |
|   | 32 | 181.5 | 14.6 | 432     | 4 | US-09-328-352-261   |
|   | 33 | 180.5 | 14.6 | 1664976 | 4 | US-08-916-421B-1    |
|   | 34 | 180.5 | 14.6 | 1664976 | 4 | US-09-692-570-1     |
|   | 35 | 178   | 14.4 | 640681  | 4 | US-09-790-988-1     |
|   | 36 | 177   | 14.3 | 1011    | 4 | US-09-107-433-1618  |
|   | 37 | 174.5 | 14.1 | 1230235 | 4 | US-09-198-452A-1    |
|   | 38 | 174.5 | 14.1 | 1230235 | 4 | US-09-438-185A-1    |
|   | 39 | 174   | 14.0 | 1026    | 3 | US-09-149-624-1     |
|   | 40 | 173   | 14.0 | 1101    | 4 | US-09-134-000C-1551 |
| C | 41 | 170   | 13.7 | 15249   | 4 | US-08-956-171E-102  |
| C | 42 | 170   | 13.7 | 15249   | 4 | US-08-781-966A-102  |
| C | 43 | 168   | 13.5 | 3064    | 3 | US-09-221-017B-194  |
| C | 44 | 159.5 | 12.9 | 42325   | 4 | US-08-311-731A-131  |
|   | 45 | 145.5 | 11.7 | 3215    | 4 | US-09-710-279-3566  |

#### ALIGNMENTS

RESULT 1

US-09-774-528-177  
Sequence 177, Application US/09774528

Patent No. 6743619

GENERAL INFORMATION:

APPLICANT: Zhou, Y. Tom

APPLICANT: Goodrich, Ryle

APPLICANT: Liu, Chenghua

APPLICANT: Asundi, Vinod

APPLICANT: Ren, Feiyan

APPLICANT: Zhang, Jie

APPLICANT: Zhao, Qing A.

APPLICANT: Yang, Yonghong

APPLICANT: Xue, Aidong J.

APPLICANT: Weinman, Tom

APPLICANT: Wang, Jian-Rui

APPLICANT: Wang, Dunrui

APPLICANT: Drmanac, Radoje T.

TITLE OF INVENTION: No. 6743619el Nucleic Acids and

FILE REFERENCE: 802 Polypeptides

CURRENT FILING DATE: 2001-01-30

NUMBER OF SEQ ID NOS: 441

SOFTWARE: pc\_fl\_genes Version 2.0

SEQ ID NO 177

LENGTH: 1416

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (205) .. (1305)

US-09-774-528-177

Alignment Scores:

Sequence 503, App  
Sequence 2841, Ap  
Sequence 2050, Ap  
Sequence 1, Appl  
Sequence 2955, Ap  
Sequence 5, Appl  
Sequence 3, Appl  
Sequence 727, App  
Sequence 3985, Ap  
Sequence 147, App  
Sequence 2, Appl  
Sequence 51, Appl  
Sequence 51, Appl  
Sequence 1072, Ap  
Sequence 214, App  
Sequence 1, Appl  
Sequence 216, Ap  
Sequence 6, Appl  
Sequence 261, App  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1618, Ap  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 102, App  
Sequence 102, App  
Sequence 794, App  
Sequence 131, App  
Sequence 3566, Ap

Pred. No.: 1,02e-157 Length: 1416  
Score: 1240.00 Matches: 239  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-09-774-528-177 (1-1416)

```
OY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
Db 586 CTGTTGGCAATTACTTCAAGGAGTTTCAGATTTCCTTCTTGAAAGCTTTGGACATA 645
OY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
Db 646 GCACCGAGTGACATGCTTGCAAGGTGGCAAGAAAGACTTTCTTAATAAATCAATCCAGAG 705
OY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
Db 706 TGTCTCCACCATGAGTGGTGGAAAGCCATGAAACATTGGCCAAACAGAAATAGATT 765
OY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80
Db 766 CATTTTGACATCAAACTCCCTTCATCATGCTAAATAATGATTTTCTTTTACTGGA 825
OY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIleGlnLys 100
Db 826 CTTCACACAGTTTACTGATTAATATATATGAAAAGAAAAGAAAGGTAATTGAGAG 885
OY 101 GlnGlnIleLeuSerSerAlaAlaAspIleAlaIleValAlaGlnHisThrMetAlaCys 120
Db 886 GGGCAAAATCTGTCTTCAAGCAGACATGCTGCCACATGACACACATGGCATGT 945
OY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140
Db 946 CATCTTGTAAGAAAGAACACATCGGGCTATTCTGTTTGTAAAGCAGAGACTTGTTACT 1005
OY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrlleArgArg 160
Db 1006 CAAATAATATGCACTACTGCTGATGCTGAGTGTCCAGATGAACTTCTATATCCGAGA 1065
OY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
Db 1066 GCTCTGGAATTTTAACAACGACACACATGCACTTGTGTGTCTCTCCAGACTA 1125
OY 181 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGlnArgLeuArgAlaGlyLeu 200
Db 1126 TGCACGTATATGCGATTATGATGATGATGATGATGATGATGATGATGATGATGATG 1185
OY 201 GlnIleLeuHisAspIleGlnGlyIleArgTyrlleProLysCysProLeuGlyValAsp 220
Db 1186 GGCATTTTACATGACATAGAGGATCGGCTATGAAACAAATGTCTCTTGAGATAGAC 1245
OY 221 IleSerLysGlnValGlyGlnAlaSerIleLysValProGlnLeuLysMetGlnIle 239
Db 1246 AATATCAAAAGAGTGGAGAGCTTCATATAAAGTACCAATTAATAATGAGAGATA 1302

RESULT 2
US-10-067-443-23
; Sequence 23, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
```

SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 23  
; LENGTH: 1526  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-067-443-23

Alignment Scores:

Pred. No.: 1,15e-157 Length: 1526  
Score: 1240.00 Matches: 239  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-067-443-23 (1-1526)

```
OY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
Db 85 CTGTTGGCAATTACTTCAAGGAGTTTCAGATTTCCTTCTTGAAAGCTTTGGACATA 144
OY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
Db 145 GCACCGAGTGACATGCTTGCAAGGTGGCAAGAAAGACTTTCTTAATAAATCAATCCAGAG 204
OY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
Db 205 TGTCTCCACCATGAGTGGTGGAAAGCCATGAAACATTGGCCAAACAGAAATAGATT 264
OY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80
Db 265 CATTTTGACATCAAACTCCCTTCATCATGCTAAATAATGATTTTCTTTTACTGGA 324
OY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIleGlnLys 100
Db 325 CTTCACACAGTTTACTGATTAATATATGAAAAGAAAAGGAGGTAATTGAGAG 384
OY 101 GlnGlnIleLeuSerSerAlaAlaAspIleAlaIleValAlaGlnHisThrMetAlaCys 120
Db 385 GGGCAAAATCTGTCTTCAAGCAGACATGCTGCCACATGACACACATGGCATGT 444
OY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140
Db 445 CATCTTGTAAGAAAGAACACATCGGGCTATTCTGTTTGTAAAGCAGAGACTTGTTACT 504
OY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrlleArgArg 160
Db 505 CAAATAATATGCACTACTGCTGATGCTGAGTGTCCAGATGAACTTCTATATCCGAGA 564
OY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
Db 565 GCTCTGGAATTTTAACAACGACACACATGCACTTGTGTGTCTCTCCAGACTA 624
OY 181 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGlnArgLeuArgAlaGlyLeu 200
Db 625 TGCACGTATATGCGATTATGATGATGATGATGATGATGATGATGATGATGATGATG 684
OY 201 GlnIleLeuHisAspIleGlnGlyIleArgTyrlleProLysCysProLeuGlyValAsp 220
Db 685 GGCATTTTACATGACATAGAGGATCGGCTATGAAACAAATGTCTCTTGAGATAGAC 744
OY 221 IleSerLysGlnValGlyGlnAlaSerIleLysValProGlnLeuLysMetGlnIle 239
Db 745 AATATCAAAAGAGTGGAGAGCTTCATATAAAGTACCAATTAATAATGAGAGATA 801

RESULT 3
US-10-067-443-1
; Sequence 1, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: SPINAL CORD, MP-1
```

FILE REFERENCE: D0073 NP  
CURRENT APPLICATION NUMBER: US/10/067,443  
CURRENT FILING DATE: 2002-02-05  
PRIOR APPLICATION NUMBER: US 60/266,518  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: US 60/282,814  
PRIOR FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1  
LENGTH: 2197  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (231)..(1472)  
US-10-067-443-1

## Alignment Scores:

|                        |           |               |      |
|------------------------|-----------|---------------|------|
| Pred. No.:             | 2,07e-157 | Length:       | 2197 |
| Score:                 | 1240.00   | Matches:      | 239  |
| Percent Similarity:    | 100.00%   | Conservative: | 0    |
| Best Local Similarity: | 100.00%   | Mismatches:   | 0    |
| Query Match:           | 100.00%   | Indels:       | 0    |
| DB:                    | 4         | Gaps:         | 0    |

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-067-443-1 (1-2197)

QY 1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
DB 756 CTGTGGCATTAAGTTCAAGAGGTTTCAGATTTCTGCTTTCGAAAGCTTTGGACATA 815  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
DB 816 GCACCAAGGTACACGCTTGACAGGTGGCAAGAGACTTTTAAATAAACATCCAGAG 875  
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGInGlyAsnArgPhe 60  
DB 876 TGCTCCACATGAGTGGTGGAGAACCATAGAACATTTGGCCAAACAGAAATGATTT 935  
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80  
DB 936 CATTTTGAATCAAAACCTCCCTTCATCATCTAAAAATTTGATTTTCTTTTACTGGA 995  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100  
DB 996 CTTCAACACGTTACTGATTAATAATATATGAAAAAGAAAAAGGAAGTATTCAGAG 1055  
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120  
DB 1056 GGGCAAAATCCGTCTTCAGCAGACATTCCTGCACAGTACAGACACAAATGSCATGT 1115  
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140  
DB 1116 CATTTTGTAAAAACACATCGGGCTATTCTGTTTGTAAAGACAGACACTTGTTACTT 1175  
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaAsnAsnPheThrIleArgArg 160  
DB 1176 CAATAATATGACAGTACTGGTTCATCTGGTGTGCGCAAGTAACTTCATATTCGCGAGA 1235  
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
DB 1236 GCTCTGCAAAATTTTAAACAAAGCAACAGTGCACCTTGTGTGTCTCTCCACAGACTA 1295  
QY 181 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
DB 1296 TGCACGTATTAATGACATTATATTCATGCAATGATGATTAAGACATAGCTGCTGCTG 1355  
QY 201 GlyIleLeuHisAspIleGluGlyIleArgGlyArgGluProLysCysProLeuGlyValAsp 220  
DB 1356 GGCATTTTACATGACATTAAGAGGCAATCCGCTATGAAACCAAAATGCTCTTGGAGTAGAC 1415  
QY 221 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239

DB 1416 ATATCAAAAGAGTTGAGAGGCTTCATTAAGTACCAATTAAATGAGATA 1472

## RESULT 4

US-10-067-443-21

Sequence 21 Application US/10067443

Patent No. 6642041

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN

TITLE OF INVENTION: SPINAL CORD, MP-1

FILE REFERENCE: D0073 NP

CURRENT APPLICATION NUMBER: US/10/067,443

CURRENT FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/266,518

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10

NUMBER OF SEQ ID NOS: 71

SOFTWARE: PatentIn version 3.0

SEQ ID NO 21

LENGTH: 1387

TYPE: DNA

ORGANISM: homo sapiens

US-10-067-443-21

## Alignment Scores:

|                        |           |               |      |
|------------------------|-----------|---------------|------|
| Pred. No.:             | 3.87e-154 | Length:       | 1387 |
| Score:                 | 1213.50   | Matches:      | 238  |
| Percent Similarity:    | 90.15%    | Conservative: | 0    |
| Best Local Similarity: | 90.15%    | Mismatches:   | 1    |
| Query Match:           | 97.86%    | Indels:       | 25   |
| DB:                    | 4         | Gaps:         | 1    |

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-067-443-21 (1-1387)

QY 1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
DB 549 CTGTGGCATTAAGTTCAAGAGGTTTCAGATTTCTGCTTTCGAAAGCTTTGGACATA 608  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
DB 609 GCACCAAGGTACATCTTGACAGGTGGCAAGAGACTTTTAAATAAACATCCAGAG 668  
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGInGlyAsnArgPhe 60  
DB 669 TGCTCCACATGAGTGGTGGAGAACCATAGACATTTGGCCAAACAGAAATGATTT 728  
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80  
DB 729 CATTTTGAATCAAAACCTCCCTTCATCATCTAAAAATTTGATTTTCTTTTACTGGA 788  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 98  
DB 81 CTTCAACACGTTACTGATTAATAATATATGAAAAAGAAAAAGGAAGTATTCAGAG 848  
QY 98 ----- 98  
DB 849 ATTAATAAGTTGAACAGATTAATAATTCCTGATTCCTGATTCCTGATTCCTGATTC 908  
QY 99 -----GluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGln 115  
DB 909 TGCAGGTATGAAGAGGCAAAATCCGTCTTTCAGCAGACATTCCTGCACAGTACAG 968  
QY 116 HisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGln 135  
DB 969 CACACATGCGATGATCTTCTGTGAAGAACAACATCGGGCTATTCTGTTTGTAAAGCAG 1028  
QY 136 ArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerArg 155  
DB 1029 AGAGACTTGTATCTCAAAATATATGACAGTACTGGTGTGATCTGGTGTGTCTCCCAAGTAC 1088  
QY 156 PheThrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCys 175

```

Db      1089 TCTATATCCGACAGAGCTCTGGAAATTTTAAACAAACGACACAGGCACTTTGTTGTCT 1148
QY      176  PPRFPRQARGLEUCYSTRASPASNGLYLLEWETLEALATRPASNGLYLLEGIUARG 195
Db      1149 CTTCCCTCCCGACAGTATGACATGATTAATGCAATTATGATTCATGACATGAAATGTAAGAA 1208
QY      196  LEUATGALGLYLEUGLYLLELEUHSAPSPILLEGUGLYLLEARGTYRCLUPPOLYSGVCS 215
Db      1209 CTACGGTGTGGCTTGGGCATTTTACATGACATPAGAGGATCCGCTATGAAACCAAAATGT 1268
QY      216  PROLENGLYVALASPILESERLYSGLYVALGLYUVALASERILEYVALPROGINLEU 235
Db      1269 CCTCTTGGATGACATATCATCAAAAGAACTTGAGAGAGCTTCATATAAAGTACCACAATTA 1328
QY      236  LYSMERGLIULE 239
Db      1329 AAAATGAGATTA 1340

RESULT 5
US-10-067-443-20
; Sequence 20, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 20
; LENGTH: 14364
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-20

Alignment Scores:
Pred. No.:      3.21e-86      Length:      14364
Score:          725.00      Matches:      186
Percent Similarity: 32.86%      Conservative: 0
Best Local Similarity: 32.86%      Mismatches: 1
Query Match:    58.47%      Indels:      380
DB:              4          Gaps:      3

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-20 (1-14364)
QY      29  VALAIAARGARGLEUSERLEULEYHSIPROGIUCYSESTHMESERGILYLYS 48
Db      11840 GGGCGAAGAAGACTTTCTTTATATAAACATCCAGAGTGTCTCCACATGAGTGTGGGAAA 11899
QY      49  AAlaIleGluHISLeuAlaLYSGInGlyASnaRgPheHISpHeASpIleLYSPropOLEu 68
Db      11900 GCCATGAGACATTTGGCCAAACAAGAAATGATTTTCATTTTGACATCAAACTCCCTTG 11959
QY      69  HISHIALALASpENCYsAsPHeSErHeThRgLYLeuGInHISValHThrAspLYSILE 88
Db      11960 CATTCAAGCTTAAAAAATTGTGATTTTCTTTACTGACCTTCAACAGCTTACTGATAAAATA 12019
QY      89  ILeMeTLYsLysGLIuLYSGInGlu----- 96
Db      12020 AATAATAAAAAGAAAGAAAGAGAAAGTATATTCTTAATTAGTAAAGTTGAACAGATAAAT 12079
QY      97  -----GlyIleGluLYSGInLI 103
Db      12080 AATTCCTGAGATTGGCTTAAAAATAGCTGCTCATTTCTGAGAGTATTTGAGAGGGGCAAT 12139
QY      103  eLeuSErSerAlaIaAspIleAlaIaThValGInHISrHrMeAlaCYsHISLeuVa 123

```

|    |       |   |       |
|----|-------|---|-------|
| Db | 12140 | CCCTCTTCAGCAGCAGACATTCCTCCACAGTACAGCAGACAGATGGCATGTCATCTTGT   | 12199 |
| OY | 123   | llveatgThHiargAlallleuPheCylysGlnatgApLeuPheProGlnAsnAs       | 143   |
| Db | 12200 | GAAGAAGACACATCGGGCTATCTGTGTTTGTATAGCAGACAGACTGTGTACCTCAAAATTA | 12259 |
| OY | 143   | naIaValIeu-----   | 146   |
| Db | 12260 | TGCAGTACTGTAGTATTATTCATTTATATAGTAATAGTTACACTTTGCCAATATGTATAC  | 12319 |
| OY | 146   | -----   | 146   |
| Db | 12320 | TTTTTCCCAAGACTTGACCTTGTGTATTAGATGGAACAGATCTTATGCTATGCTAG      | 12379 |
| OY | 146   | -----   | 146   |
| Db | 12380 | CCCTGACAGTATGAAATTATGACAGATAGGAAGAACTAAACGCCATTTCTGTACTAGTT   | 12439 |
| OY | 146   | -----   | 146   |
| Db | 12440 | TGTAAGCTTTATGGGACAGCTGTATAGCTTATAGGCACATAGTCTAAATTTGCATCTT    | 12499 |
| OY | 146   | -----   | 146   |
| Db | 12500 | CTTGTGGATTTAAAGAGGCTTACAAATTAAGAAAGTAATGCACTACGTCTATCACT      | 12559 |
| OY | 146   | -----   | 146   |
| Db | 12560 | ATTTTATGAAAAATAGGTGATTTCTTCTCATCTTGATGAATCCCTTGTGTGTGTGT      | 12619 |
| OY | 146   | -----   | 146   |
| Db | 12620 | TTTTTATATAGCCAGTCAATTATAGCAGTGGAGGTGTATTCACACTTGTGTCAGACT     | 12679 |
| OY | 146   | -----   | 146   |
| Db | 12680 | AATGTGATTAAGTCTTGATTAATCCACTATATTGTACACGCCAAAATCCCTTAATTGTC   | 12739 |
| OY | 146   | -----   | 146   |
| Db | 12740 | CTTAAAGCCTTGACAAACATCCTGTTTAACTGTATCTTAACCTTATTCATTTAAAT      | 12799 |
| OY | 146   | -----   | 146   |
| Db | 12800 | TATTAATAAGTGGGAAAATGTTTAAATGTAGTAATCAATAGATGAATTTTACATG       | 12859 |
| OY | 146   | -----   | 146   |
| Db | 12860 | ATATCAAGATTAATTTTTCAGAGTTATGTAGTAAATGCACAAAATATATAAAATTTTC    | 12919 |
| OY | 146   | -----   | 146   |
| Db | 12920 | AGGCTTAATAATGTACTATGATTTGAATATATTAATTAATAATTTTATAGTGAAGGT     | 12979 |
| OY | 146   | -----   | 146   |
| Db | 12980 | TGGAAGAAAATATACAAAATGCTAGTAATGTTGTATGCTATATAGAAATTAATTAAT     | 13039 |
| OY | 146   | -----   | 146   |
| Db | 13040 | TTTTTCTTCCAAATTTTATTAATACATAGATATGTCATCTGCCCATTAACCATCTCAAAAT | 13099 |
| OY | 146   | -----   | 146   |
| Db | 13100 | GGGATAGTTATTAATGTTTAAATGCTGATATTTTCTCCAGGTTTAATTAGCAGCTGTGT   | 13159 |
| OY | 147   | -----ValIaSerGlyGcl   | 151   |
| Db | 13160 | TCAATATCCATATATGATAGTATTTTGTGTTTCTCAATTCCTTCAGGTTGCATCTGTGTG  | 13219 |
| OY | 151   | YValIaSerAspPheTyrIleArgArgAlaIeuGluIleuThrAsnAlaThrGlnCly    | 171   |

Db 13220 TGTGCAAGTAAGTCTTATATCCGACAGCTCTGAAATTTTAAACCAACGACAGCTG 13279  
Qy 171 sThrLeuLeuCyProProProProLeuCyThraAspAsnGlyIleMetIleAlaTrp-- 190  
Db 13280 CACTTTGTTGTCTCTCTCCAGACTATGACATGATATGATGATTTGATTCATGTA 13339  
Qy 190 ----- 190  
Db 13340 AGCCACAGATATACGTCTTCACTCACTAATATATTAATTCATTTTATCAT 13399  
Qy 190 ----- 190  
Db 13400 ACTAAGCTTCTTCTTCAGATCTTGAGCTATGATTTATTTATGCTTTATTTA 13459  
Qy 191 --AsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleA 210  
Db 13460 GGAATGATATGAAAGCTACAGTCTGCTGGCTTGGCAATTTTACATGACATATAGAGCATCC 13519  
Qy 210 rGTYrGluProlys 214  
Db 13520 GCTATGACCAAG 13533  
RESULT 6  
US-09-596-002-38  
Sequence 38, Application US/09596002  
Patent No. 6632636  
GENERAL INFORMATION:  
APPLICANT: Lagace, Robert, E.  
APPLICANT: Patterson, Chandra  
APPLICANT: Berg, Klm, E.  
TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME  
FILE REFERENCE: PM-0008-4 US  
CURRENT APPLICATION NUMBER: US/09/596,002  
PRIOR FILING DATE: 2000-06-16  
PRIOR FILING DATE: 1999-06-18  
NUMBER OF SEQ ID NOS: 41  
SOFTWARE: PERL Program  
SEQ ID NO 38  
LENGTH: 94750  
TYPE: DNA  
ORGANISM: M. catarrhalis  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: incyte template ID No. 6632636 38  
PUBLICATION INFORMATION:  
US-09-596-002-38  
Alignment Scores:  
Pred. No.: 3,02e-20 Length: 94750  
Score: 248.50 Matches: 73  
Percent Similarity: 48.67% Conservative: 37  
Best Local Similarity: 32.30% Mismatches: 90  
Query Match: 20.04% Indels: 27  
Gaps: 7  
DB: 4  
US-10-649-273-2\_COPY\_176\_414 (1-239) x US-09-596-002-38 (1-94750)  
Qy 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLeuSerLeuAspIle 20  
Db 24305 ATGCTGTGCGTCCGATGCTGTGGCGGTATCATGATATTCGGAGATCTATCATGAT 24364  
Qy 21 AlaProGlyAspMetLeuAspLeuValAlaArgArgLeuSerLeuIleLeuHisProGlu 40  
Db 24365 GCGGTGGTGAATGCTTGAATAAACGCAAAATGCTCAATG---CCCTATCT--- 24418  
Qy 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaLeuGlnGlyAsnArgPhe 60  
Db 24419 -----GGTGGCGCCAAATATCGAAAAATTTAGCCAAAAACGCAAAACCCACAC 24463  
Qy 61 HisPheAspIleLeuSProProLeuHisIleAlaLeuAsnCyAspAspPheSerTherGly 80  
Db 24464 GCCTATAGCTGCCAAGACCATCATGACAT---AAAGGGCTGAGATTTCCTTCAGTGC 24520

Qy 81 LeuGlnHisValThrAspLeuIleMetLeuSylsGluLeuGluGlyIleGluLeu 100  
Db 24521 ATGAAACCGCATTCATTAATCTCAAAAGACACACAAACGCCAAACGACCC--- 24577  
Qy 101 GlyGlnIleLeuSerSerAlaIleAspIleAlaIleThrValGlnHisThrMetAlaCys 120  
Db 24578 -----GCCACAGAGACATCGCGCAAGCTTTGAGTATGCGGTGGAT 24625  
Qy 121 HisLeuValIleArgThrHisArgAlaIleLeuPheCysLeuGlnArgAspLeuPro 140  
Db 24626 ACTTGGTCAAAAATGACACCAACACTACAGATGACAGGATTCGCCAG----- 24676  
Qy 141 GlnAsnAlaValLeuValAlaSerGlyValAlaSerAspPheThrIleArgArg 160  
Db 24677 -----CTGGTGTCCAGAGGGGCGTCTGCCAATCAGATGCTACGCCCC 24721  
Qy 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCyProProArgLeu 180  
Db 24722 ACCCTGACCGACGCTCCGCAATCGATGCGGTGCTACTATGCCCCGACGACTA 24781  
Qy 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
Db 24782 TGCACGATTAATGGTCCGATGATCGCTATGCTGCTTTTGTGCGCTCAGCTGTGACAG 24841  
Qy 201 GlyIleLeuHisAspIleGluGlyIleArgTyrgluProlysCyProLeuGlyValAsp 220  
Db 24842 TCG-----GATGATTTGGC-GTTCGCGTATTC-----CCGATGGATATGAC 24885  
Qy 221 IleSerLysGluValGly 226  
Db 24886 GACGCTTGCGATCGGC 24903  
RESULT 7  
US-09-540-236-806  
Sequence 806, Application US/09540236  
Patent No. 6673910  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS  
FILE REFERENCE: 2709.2005-001  
CURRENT APPLICATION NUMBER: US/09/540,236  
PRIOR FILING DATE: 2000-04-04  
NUMBER OF SEQ ID NOS: 3840  
SEQ ID NO 806  
LENGTH: 1053  
TYPE: DNA  
ORGANISM: M. catarrhalis  
US-09-540-236-806  
Alignment Scores:  
Pred. No.: 3,22e-23 Length: 1053  
Score: 247.00 Matches: 70  
Percent Similarity: 46.85% Conservative: 34  
Best Local Similarity: 31.53% Mismatches: 80  
Query Match: 19.92% Indels: 38  
Gaps: 6  
DB: 4  
US-10-649-273-2\_COPY\_176\_414 (1-239) x US-09-540-236-806 (1-1053)  
Qy 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLeuSerLeuAspIle 20  
Db 445 ATGCTGTGCGTCCGATGCTGTGGCGGTATCATGATATTCGGAGATCTATCATGAT 504  
Qy 21 AlaProGlyAspMetLeuAspLeuValAlaArgArgLeuSerLeuIleLeuHisProGlu 40  
Db 505 GCGGTGGGGAATGCTTGAATAAACGCAAAATGCTCAATG---CCCTATCT--- 558  
Qy 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaLeuGlnGlyAsnArgPhe 60  
Db 559 -----GCTGTCCAAATATCGAAAAATTTAGCCAAAAACGCAAAACCCACAC 603

[illegible]

US-09-252-991A-884/c  
 Sequence 884, Application US/09252991A  
 Patent No. 6551795  
 GENERAL INFORMATION:  
 APPLICANT: Marc J. Rubenfield et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 107196.136  
 CURRENT APPLICATION NUMBER: US/09/252,991A  
 CURRENT FILING DATE: 1999-02-18  
 PRIOR APPLICATION NUMBER: US 60/074,788  
 PRIOR FILING DATE: 1998-02-18  
 PRIOR APPLICATION NUMBER: US 60/094,190  
 PRIOR FILING DATE: 1998-07-27  
 NUMBER OF SEQ ID NOS: 33142  
 SEQ ID NO 884  
 LENGTH: 1059  
 TYPE: DNA  
 ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-884

|                                      |                             |
|--------------------------------------|-----------------------------|
| Alignment Scores:                    |                             |
| Pred. No.:                           | 4,81e-21                    |
| Score:                               | 231.00                      |
| Percent Similarity:                  | 49.54%                      |
| Best Local Similarity:               | 33.80%                      |
| Query Match:                         | 18,63%                      |
| DB:                                  | 4                           |
| US-10-649-273-2_COPY_176_414 (1-229) | US-09-252-991A-884 (1-1059) |
| Length:                              | 1059                        |
| Matches:                             | 73                          |
| Conservative:                        | 34                          |
| Mismatches:                          | 87                          |
| Indels:                              | 22                          |
| Gaps:                                | 8                           |

QY 2 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyValSerLeuAspIleAla 21  
||| ||| :::: ||||| ::||| |||  
Db 621 TTGGTGGCGGGTGGACGGTATCGGGCCGTCACAGTTCTTGGCGAATCGGTGGACGATGCC 562  
QY 22 ProGlyAspMetLeuAspIlyValAlaArgArgLeuSerLeuIleValHisProGlyCys 41

Db 561 GCCCGGAAAGCCTTGACAAAGACCGGCAAGCATGATGGCGTGG---GGCATATGCC-----511111  
 GY 42 SerThmetSerGlyGlyValaIleGluHisLeuAlaGlyGlnGlyAsnArgPheHis 61  
 Db 510 -----GGTGGTCCGGAATCGCCCGCTGGCGGAGCGCGGCACTCTGGCGCG 463  
 GY 62 PheAspIleValProProLeuHisHisAlaValAsnGlyAspPheSerPheThrGlyLeu 81  
 Db 462 TTTCGTGTTCCGCGCGCGCATGACCGATCGCCCGCGCTGGACTTTCAGCTTTCACGCGGCTC 403  
 GY 82 GlnHisValThrAspValIleIleMetGlyValGlyGlnGluGlyIleGluGlyGly 101  
 Db 402 AAGACCTTTACCTGACAC--ACCTGGCAGCGCTTGGCTGAGAGCGGCGACGACGAG 346  
 GY 102 GlnIleuSerSerSerAlaAlaAspIleAlaIleValGlnHisThrMetAlaCysHis 121  
 Db 345 CAG-----ACCGCGTGCAGACATCGCTTCGGCGCTTCCAGACCGCGGTCGAGACC 295  
 GY 122 LeuValIleArgThrHisArgAlaIleLeuPheCysGlyGlnArgAspLeuProGln 141  
 Db 294 CTGCTGATCAAGTGGCGCTGGCGCCCTGG-----AACGACAGCGGCGCTG-----AAG 250  
 GY 142 AsnAsnAlaValLeuValAlaSerGlyValAlaSerAspPheThrIleArgAla 161  
 Db 249 AAC-----CTGGTGATCGCCGCGGATGTGACAGGCAACACGACGCTGCGACGCGC 199  
 GY 162 LeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaGluCys 181  
 Db 198 CTGGAAAGATGCTCGCGCAATGAAAGGCGGAGAGTGTCTCAAGCCCGCGCGCTTCTGC 139  
 GY 182 ThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGly 201  
 Db 138 ACCGACATGCGGCGCATGATGCGCTAAGCGCGGCTGCAAGCGCTGCTGCGCGGC-----85  
 GY 202 IleLeuHisAspIleGluGlyIleArgThrGlyProGlyCysProLeu 217  
 Db 84 ---CAGCATGAGCGCCGCGCATGACAGCTTCAAGCGCGCGCTGGCCATG 40

RESULT 9  
 US-09-252-991A-801  
 : Sequence 801, Application US/09252991A  
 : Patent No. 6551795  
 : GENERAL INFORMATION:  
 : APPLICANT: Marc J. Rubenfield et al.  
 : TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 : TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 : FILE REFERENCE: 107196.136  
 : CURRENT APPLICATION NUMBER: US/09/252,991A  
 : PRIORITY FILING DATE: 1999-02-18  
 : PRIOR APPLICATION NUMBER: US 60/074,788  
 : PRIOR FILING DATE: 1998-02-18  
 : PRIOR APPLICATION NUMBER: US 60/094,190  
 : PRIOR FILING DATE: 1998-07-27  
 : NUMBER OF SEQ ID NOS: 33142  
 : SEQ ID NO 801  
 : LENGTH: 1206  
 : TYPE: DNA  
 : ORGANISM: Pseudomonas aeruginosa  
 : US-09-252-991A-801

|                        |          |                  |
|------------------------|----------|------------------|
| Alignment Score:       |          |                  |
| Pred. No.:             | 5,956-21 | Length: 12066    |
| Score:                 | 231.00   | Matches: 73      |
| Percent Similarity:    | 49.54    | Conservative: 34 |
| Best Local Similarity: | 33.80    | Mismatches: 87   |
| Query Match:           | 18.63    | Indels: 22       |
| DB:                    | 4        | Gaps: 8          |

US-10-649-273-2\_COPY\_176\_414 (1-229) x US-09-252-991A-801 (1-1206)  
Qy 2 LeuAlaLeuValGInGlyValSerAspPheLeuLeuGluYLySerLeuAspIleLea 21  
||| ||| |::: ::| |||||||:::||:::|| ||



```

Db 604 TTGTCGGGGTGACGGATCGCCGCGCTACAGCTTGGCGAATCGTCGACGATGCC 663
Qy 22 ProGlyAspMetLeuAspLysValAlaArgLysSerLeuIleuLysHisProGlyCys 41
Db 664 GCGGGGAAGCCTTCGACAAAGACCGCCAGCTGATCGGCTG--GGCTATCCC----- 714
Qy 42 SerThiMetSerGlyGlyLysAlaAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHis 61
Db 715 -----GGTGTCCGGAATCGCCGCGCTGCGGAGCGCGGCACTCCGCGCGC 762
Qy 62 PheAspIleLysProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeu 81
Db 763 TTCGTGTTCCCGCGCGCATCGACCGATCGCCCGGCTTCACATTCACGCTTCAGCGGCTC 822
Qy 82 GlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIleGluLysGly 101
Db 823 AAGACCTTACCTTGAC--ACCTGGACAGCTTGCGCTCGAGCGCGGACAGACGCGAG 879
Qy 102 GlnIleLysSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHis 121
Db 880 CAG-----ACCGCTCGGACATCGCCCTGCGCTTCAGACCGCGGCTGCGAGACC 930
Qy 122 LeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGln 141
Db 931 CTGCTGATCAGTGCCTGCGCGCTTG-----AAGCAGACCGGCTG-----AAG 975
Qy 142 AsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPhePheThrIleArgAla 161
Db 976 AAC-----CTGTGTATCGCGCGGCGGTGTCAGCCCAACAGCGGCTGCGAGCGGC 1026
Qy 162 LeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCys 181
Db 1027 CTGGAAGAAGATGTCGCGCAATGAAGAGGCGAGGTGTTACGCCCGCGCGCTTCTGC 1086
Qy 182 ThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGly 201
Db 1087 ACCGAAATGGCGCGATGATCGCTACGCCGCGTCCAGCGCGCTGCTGCGCGCGC----- 1140
Qy 202 IleLeuHisAspIleGluGlyIleArgGlyLeuGluProLysCysProLeu 217
Db 1141 ---CAGCATGACGCGCGCGCATCGAGCTCAGCGCGCTGCGCGCATG 1185

```

RESULT 10  
US-09-557-884-1/c  
Sequence 1, Application US/09557884  
Patent No. 6506581

## GENERAL INFORMATION:

APPLICANT: Fleischmann et al.  
TITLE OF INVENTION: The Nucleotide sequence of  
the Haemophilus influenzae Rd Genome, Fragments  
Thereof, and Uses Thereof

## NUMBER OF SEQUENCES: 1

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: MD  
COUNTRY: USA  
ZIP: 20850

## COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette  
COMPUTER: Dell Pentium  
OPERATING SYSTEM: MS DOS V6.22  
SOFTWARE: ASCII Text

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/557,884  
FILING DATE: 25-Apr-2000  
CLASSIFICATION: <Unknown>

## PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 08/476,102  
FILING DATE: JUN-5-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Michelle S. Marks

```

REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186F3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

```

```

Alignment Scores:
Pred. No.: 7,86-15 Length: 1830121
Score: 224.00 Matches: 67
Percent Similarity: 47.50% Conservative: 28
Best Local Similarity: 33.50% Mismatches: 83
Query Match: 18.06% Indels: 22
DB: 4 Gaps: 6

```

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-09-557-884-1 (1-1830121)

```

Qy 2 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 21
Db 552791 TTACTGGGTGTCGATGGTGATGGAATAATGAGATGATGAGAAATATTGATGATGCT 552732
Qy 22 ProGlyAspMetLeuAspLysValAlaArgLysSerLeuIleLysHisProGlyCys 41
Db 552731 GCTGGCAAGCCTTGATTAATAACGAAATTAATCTTGACATTA--GATTAATCCA----- 552681
Qy 42 SerThiMetSerGlyGlyLysAlaAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHis 61
Db 552680 -----GGTGGCGCGGCACTTCTGCTTAGCGGAATAAGGTAACCGCAATGCT 552633
Qy 62 PheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeu 81
Db 552632 TTCACATTCACAGCTCAATACAGATGTCGAGCCCTTGATTTTTCGTGTTA 552573
Qy 82 GlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIleGlu 99
Db 552572 AAAACATTGCGCGCAATATACGTAATCAACAAATTAATAAAGAGGCGCAACTGATAG 552513
Qy 100 LysGlyGlnIleLysSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAla 119
Db 552512 CAA-----ACTAAAGCAGATATGCTTATGCTTTCACAGATGCGGTGCTG 552468
Qy 120 CysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeu 139
Db 552467 GATACCTCTTGCC-----ATTAAATGTAAAG--CGTGCATTGAAA 552432
Qy 140 ProGlnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArg 159
Db 552431 GAAACAGCTATAAACGTTTATGATTCGCGGAGGCGTGAAGCGCAATAAATAAATCCGA 552372
Qy 160 ArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArg 179
Db 552371 GAAACGCTTGCGCATATATGCAAAATTTAGGTGCGAAGTGTATTATCTCAACTCAA 552312
Qy 180 LeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGly 199
Db 552311 TTTGTATACGATTAATGTGCGATGATGCTTACACAGAGTTTATTATGTTAAACAGAGT 552252

```

RESULT 11  
US-09-643-990A-1/c  
Sequence 1, Application US/09643990A  
Patent No. 6528289

## GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann  
Mark D. Adams  
Owen White  
Hamilton O. Smith



Db 685 GGGCAGGCGCTG-----GGGATTTGTGCGCTCTTCCAGAGAGCGCGCTGCGAC 735  
Qy 121 HisLeuValIysArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuPro 140  
Db 736 GTGCTGTCAAGAG-----CTGTGCGCGCGCGCGCGCGCTTG----- 774  
Qy 141 GlnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPhenylIleArgArg 160  
Db 775 ---GGCCACAAGAGATTGCTGTGCGCGCGCTGCGCGAATCGCGCGCTGCGGCA 831  
Qy 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180  
Db 832 CTGTGTCAAGCGCGAGCGCGAGCGCGGCTTGAACATGTTCTGCGCGCGCTGCGCTG 891  
Qy 181 CysThrAsnAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
Db 892 TGCACGACATGCGCGCATGATTGCGGTGCGCGGAGTATGAGCGGTACCGCGCGCTG 951  
  
RESULT 13  
US-09-902-540-503/c  
Sequence 503, Application US/09902540  
Patent No. 6833447  
GENERAL INFORMATION:  
APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Wiegand, Roger C.  
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-10(15849)B  
CURRENT APPLICATION NUMBER: US/09/902,540  
PRIOR FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/217,883  
PRIOR FILING DATE: 2000-07-10  
NUMBER OF SEQ ID NOS: 16825  
SEQ ID NO 503  
LENGTH: 2582  
TYPE: DNA  
ORGANISM: Myxococcus xanthus  
US-09-902-540-503  
  
Alignment Scores:  
Pred. No.: 3,98e-19 Length: 2582  
Score: 221.50 Matches: 70  
Percent Similarity: 47.50% Conservative: 25  
Best Local Similarity: 35.00% Mismatches: 74  
Query Match: 17.86% Indels: 31  
Gaps: 8  
US-10-649-273-2\_COPY\_176\_414 (1-239) x US-09-902-540-503 (1-2582)  
Qy 5 ValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyAsp 24  
Db 1096 GTCCAGGCGCTCAAGCGGAGTACCGGCTGTGGGAGACAGCGCGACAGCGCGCGCGAG 1037  
Qy 25 MetLeuAspLysValAlaIleArgArgLeuSerLeuIleLysIleProGluCysSerThrMet 44  
Db 1036 GCATATACAAAGCCGCTCGCATCTCGCGCTG---CCGATATCCG----- 995  
Qy 45 SerGlyGlyLysAlaIleGlnIleLeuAlaLysGlnGlyAsn-----ArgPhe 60  
Db 994 ---GGTGGGAGCGCCATGACAGTGGCGAGAGGGGAAACCGGAGCGCATCCGCTTC 938  
Qy 61 HisPheAspIleLysProProLeuHisIleValLysAsnCysAspPheSerPheThrGly 80  
Db 937 -----CCCGCGCGCGCTCGCGCGAGCAACTTCGATCTCTTCGCGG 893  
Qy 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlyGluGluGlyIleGluLys 100  
Db 892 TTGAAG-----ACGCGCGTCTGCACACCGTGCAGAGCAAGCGCGTGCAG 845  
Qy 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaLys 120  
Db 844 GGGCAGGCGCTG-----GCGGATTGTGCGCGCTCTTCAGAGAGCGCGTGCAGAC 794

Qy 121 HisLeuValIysArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuPro 140  
Db 793 GTGCTGTCAAGAG-----CTGTGCGCGCGCGCGCGCGCTTG----- 755  
Qy 141 GlnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPhenylIleArgArg 160  
Db 754 ---GGCCACAAGAGATTGCTGTGCGCGCGCTGCGCGAATCGCGCGCTGCGGCA 698  
Qy 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180  
Db 697 CTGTGTCAAGCGCGAGCGCGAGCGGAGGCTTGAACATGTTCTGCGCGCGCTGCGCTG 638  
Qy 181 CysThrAsnAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
Db 637 TGCACGACATGCGCGCATGATTGCGGTGCGCGGAGTATGAGCGGTACCGCGCGCTG 578  
  
RESULT 14  
US-09-543-681A-2341  
Sequence 2341, Application US/09543681A  
Patent No. 6605709  
GENERAL INFORMATION:  
APPLICANT: GARY BRETON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
FILE REFERENCE: 2709.1002-001  
CURRENT APPLICATION NUMBER: US/09/543,681A  
PRIOR FILING DATE: 2000-04-05  
PRIOR APPLICATION NUMBER: US 60/128,706  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 8344  
SEQ ID NO 2341  
LENGTH: 1074  
TYPE: DNA  
ORGANISM: Proteus mirabilis  
US-09-543-681A-2341  
  
Alignment Scores:  
Pred. No.: 1.86e-18 Length: 1074  
Score: 212.00 Matches: 66  
Percent Similarity: 44.95% Conservative: 32  
Best Local Similarity: 30.28% Mismatches: 90  
Query Match: 17.10% Indels: 30  
Gaps: 7  
US-10-649-273-2\_COPY\_176\_414 (1-239) x US-09-543-681A-2341 (1-1074)  
Qy 2 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 21  
Db 475 TTAATTAGTGTAAACGAGGATTGGGAAATATACCTCTTAGTGAGTCTATGATGATGCT 534  
Qy 22 ProGlyAspMetLeuAspLysValAlaIleArgArgLeuSerLeuIleLysHisProGluCys 41  
Db 535 GCTGTGAAGCATTTGATTAACAGCAAGCATATGCGGCTT---GATATATCC----- 585  
Qy 42 SerThrMetSerGlyGlyValAlaIleGlnIleLeuAlaLysGlnGlyAsnArgPheHis 61  
Db 586 -----GGGGCGCTGTATTATCAAAATAGCAACAAGTGTAGAGAGAGCT 633  
Qy 62 PheAspIleLysProProLeuHisIleValLysAsnCysAspPheSerPheThrGlyLeu 81  
Db 634 TTGTGTTTCTCGTCCATGACAGACAGACCGGAGCTGACCTTATGTTTCTCTGCTTA 693  
Qy 82 GlnHisValThrAspLysIleIleMetLysLysGlyGluGluGlyIleGluLysGly 101  
Db 694 AAACCTTGGCGCGCATATGATTTGTCAAAGCATGATTCAGAG----- 738  
Qy 102 GlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHis 121  
Db 739 -----CAAACTGACAGATATGTCGCTGCTTTGAGAGANCCCTAGTACTACT 789  
Qy 122 LeuValIysArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuProGln 141

Db 790 TTGGCAATAAATGCTGACGA-----TTAGACAA 822  
Qy 142 AsnAsn---AlaValLeuValAlaSerGlyGlyValAlaSerAsnPhetYrIleArg 160  
Db 823 ACAGCGCTTAAGCGCTTAGATGATGCTGGGGCGCTTAAGCTTAACCGGCC 882  
Qy 161 AlaleuGluIleleuThrAsnAlaThrGlnCysThrIleuLeuCysProProProArgLeu 180  
Db 883 AAAATGCGCATGTATATGAAACAACCTCGAGGGGAAGTGTATATGCTCGCCCTGAGCTTA 942  
Qy 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
Db 943 TGTACCGAATAATGCTGATGCTGCTGGCGGGAGTGAATCCGTTTAAAGGTGAACC 1002  
Qy 201 ---GlyIleleuHisAspIleGluGlyIleArgTyrgIuProIysCysProleu 217  
Db 1003 GAGGGGCGCTTTA-----GGGGTGAACGTGAGACACGCTTGGCCCTTTA 1044

## RESULT 15

US-09-489-039A-2050  
; Sequence 2050, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489, 039A  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 2050  
; LENGTH: 1032  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-2050

## Alignment Scores:

| Pred. No.:             | 1.55e-17 | Length:       | 1032 |
|------------------------|----------|---------------|------|
| Score:                 | 205.00   | Matches:      | 66   |
| Percent Similarity:    | 43.58%   | Conservative: | 29   |
| Best Local Similarity: | 30.28%   | Mismatches:   | 93   |
| Query Match:           | 16.53%   | Indels:       | 30   |
| DB:                    | 4        | Gaps:         | 7    |

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-09-489-039A-2050 (1-1032)

Qy 2 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyIysSerLeuAspIleAla 21  
Db 442 TTGATTAGCGGTACCGGATTTGTCAGTATGAATCTGCTGGCGAGTGAAGATGCG 501  
Qy 22 ProGlyAspMetLeuAspIysValAlaArgArgLeuSerLeuIleYshisProGluCys 41  
Db 502 GCGGGGAGCGCTTATATAGACGCGAAGCTCTGCGACTG--GATTATCCC----- 552  
Qy 42 SerThrMetSerGlyIysAlaIleGluHisLeuAlaIysGlnGlyAsnArgPheHis 61  
Db 553 -----GCGCGGCGGATGCTGCGAAATGCGTCCAGGCGCACCGAAGCGCGC 600  
Qy 62 PheAspIleIysProProLeuHisIshAlaIysAsnCysAspPheSerPheThrGlyLeu 81  
Db 601 TTTGTTTCCCGGCGCGATGACCGATCCCGGGCTGACATTCACTTCCGCGCTG 660  
Qy 82 GlnHisValThrAspIysIleIleMetIysGluGluGluGlyIleGluIysGly 101  
Db 661 AAGACCTTCGCGCAACACATTCCGACGACGCGACGACGATGAG----- 705  
Qy 102 GlnIleLeuSerSerAlaAspIleAlaIleAlaThrValGlnHisIshMetAlaCysHis 121  
Db 706 -----CAACCGCGCGCATTCGCGCGCGCTTTAGAGATGCGGTCGATACG 756  
Qy 122 LeuValIysArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuLeuProGln 141

Db 757 CTGATGATTAAATGCTGCGCGCG-----CTGAGACAA 789  
Qy 142 AsnAsn---AlaValLeuValAlaSerGlyGlyValAlaSerAsnPhetYrIleArg 160  
Db 790 ACCGCGCTTAAGCGCTTAGATGATGCTGGGGCGCTTAAGCTTAACCGGCC 849  
Qy 161 AlaleuGluIleleuThrAsnAlaThrGlnCysThrIleuLeuCysProProProArgLeu 180  
Db 850 AAGCTGCGGAGATGATGCAAAAACGCGCGCGAGGTGTTCTACGCCCGCTGAGTTTC 909  
Qy 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
Db 910 TGTACTGACAAACGCGCGGATGATGCTTACGCGCGCATGTGCGTTCGCAAAACGCGCGC 969  
Qy 201 GlyIleleuHisAspIleGlu---GlyIleArgTyrgIuProIysCysProleu 217  
Db 970 -----AAAGCGAGCTGCGCGTGAACGCTTGGCCCGCGCTGCGCTG 1011

Search completed: June 17, 2005, 02:30:33  
Job time : 984.721 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: June 16, 2005, 23:18:53 ; Search time 1623.64 Seconds

(without alignments)  
913.751 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414

Perfect score: 1240  
Sequence: 1 L1ALVQGVSDFLILGKSLDI.....DISKEVGEASIKVPOLKMEI 239

Scoring table:

|                           |  |
|---------------------------|--|
| BLOSUM62                  |  |
| Xgapop 10.0 , Xgapext 0.5 |  |
| Xgapop 10.0 , Xgapext 0.5 |  |
| Xgapop 6.0 , Xgapext 7.0  |  |
| Delop 6.0 , Delext 7.0    |  |

Searched: 6054689 segs, 3103772919 residues

Total number of hits satisfying chosen parameters: 12109378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame\_plus.model -DEV=slp  
-O=/cgn2\_1/USPT.spool.p/US10649273/runat\_15062005\_111418\_6138/app.query.fasta\_1.1429  
-DB=Published Applications NA -QEXT=fastcap -SUFFIX=p2n.rmp -MINMATCH=0.1  
-LOOPL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu2  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USR=US10649273\_@CGN\_1\_1\_1034\_@runat\_15062005\_111418\_6138  
-NCPU=6 -ICPU=3 -NO MMAP -LAKEQUERRY -NEG SCORES=0 -WMIT -DSBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:\*

|     |  |
|-----|--|
| 1:  | /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*  |
| 2:  | /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*   |
| 3:  | /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*  |
| 4:  | /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*  |
| 5:  | /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*  |
| 6:  | /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:* |
| 7:  | /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*  |
| 8:  | /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*  |
| 9:  | /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:* |
| 10: | /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:* |
| 11: | /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:* |
| 12: | /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*  |
| 13: | /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:* |
| 14: | /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:* |
| 15: | /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:* |
| 16: | /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:* |
| 17: | /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:* |
| 18: | /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:* |
| 19: | /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:* |
| 20: | /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:* |
| 21: | /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:* |
| 22: | /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*  |
| 23: | /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:* |
| 24: | /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*  |
| 25: | /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*  |
| 26: | /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*  |

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length  | DB ID | Description          |
|------------|--------|-------------|---------|-------|----------------------|
| 1          | 1240   | 100.0       | 1416    | 17    | US-10-120-988-177    |
| 2          | 1240   | 100.0       | 1526    | 17    | US-10-067-443-23     |
| 3          | 1240   | 100.0       | 1526    | 18    | US-10-649-273-23     |
| 4          | 1240   | 100.0       | 1526    | 18    | US-10-651-722-23     |
| 5          | 1240   | 100.0       | 2197    | 14    | US-10-067-443-1      |
| 6          | 1240   | 100.0       | 2197    | 18    | US-10-649-273-1      |
| 7          | 1240   | 100.0       | 2197    | 18    | US-10-651-722-1      |
| 8          | 1240   | 100.0       | 2572    | 21    | US-10-480-988-36     |
| 9          | 1213.5 | 97.9        | 1387    | 14    | US-10-067-443-21     |
| 10         | 1213.5 | 97.9        | 1387    | 18    | US-10-649-273-21     |
| 11         | 1213.5 | 97.9        | 1387    | 18    | US-10-651-722-21     |
| 12         | 1213.5 | 97.9        | 1387    | 21    | US-10-887-553A-1047  |
| 13         | 1203   | 97.0        | 1245    | 14    | US-10-012-140-6      |
| 14         | 1203   | 97.0        | 1820    | 14    | US-10-012-140-4      |
| 15         | 1059   | 85.4        | 2208    | 17    | US-10-094-749-400    |
| 16         | 1059   | 85.4        | 2890    | 20    | US-10-723-860-7447   |
| 17         | 725    | 58.5        | 14364   | 14    | US-10-067-443-20     |
| 18         | 725    | 58.5        | 14364   | 18    | US-10-649-273-20     |
| 19         | 725    | 58.5        | 14364   | 18    | US-10-651-722-20     |
| 20         | 468    | 37.7        | 371     | 19    | US-10-430-201-3118   |
| 21         | 468    | 37.7        | 371     | 19    | US-10-430-201-3119   |
| 22         | 337    | 27.2        | 1917    | 18    | US-10-424-599-66417  |
| 23         | 270    | 21.8        | 1628    | 19    | US-10-437-963-11249  |
| 24         | 257    | 20.7        | 1146    | 17    | US-10-282-122A-14674 |
| 25         | 248.5  | 20.0        | 1000    | 19    | US-10-143-561-50     |
| 26         | 248.5  | 20.0        | 1044    | 17    | US-10-282-122A-26972 |
| 27         | 248.5  | 20.0        | 94750   | 18    | US-10-672-787-38     |
| 28         | 247    | 19.9        | 936     | 17    | US-10-282-122A-8315  |
| 29         | 241    | 19.4        | 1032    | 17    | US-10-282-122A-31043 |
| 30         | 239    | 19.3        | 756     | 14    | US-10-081-051-8      |
| 31         | 239    | 19.3        | 4360    | 14    | US-10-081-051-2      |
| 32         | 231    | 18.6        | 1026    | 9     | US-09-815-242-7701   |
| 33         | 231    | 18.6        | 1026    | 17    | US-10-282-122A-30016 |
| 34         | 224    | 18.1        | 1029    | 17    | US-09-815-242-6946   |
| 35         | 224    | 18.1        | 1029    | 17    | US-10-282-122A-22020 |
| 36         | 224    | 18.1        | 1830121 | 17    | US-10-329-670-1      |
| 37         | 224    | 18.1        | 1830121 | 20    | US-10-158-865-1      |
| 38         | 213    | 17.2        | 1014    | 9     | US-09-815-242-9682   |
| 39         | 213    | 17.2        | 1014    | 9     | US-10-282-122A-39301 |
| 40         | 212    | 17.1        | 1020    | 17    | US-10-282-122A-32554 |
| 41         | 210    | 16.9        | 1014    | 17    | US-10-282-122A-41977 |
| 42         | 208    | 16.8        | 1007    | 17    | US-10-282-122A-19920 |
| 43         | 208    | 16.8        | 1728    | 19    | US-10-437-963-9137   |
| 44         | 207.5  | 16.7        | 1023    | 17    | US-10-282-122A-31809 |
| 45         | 206    | 16.6        | 1014    | 9     | US-09-815-242-6207   |

#### ALIGNMENTS

RESULT 1  
US-10-120-988-177  
Sequence 177, Application US/10120988  
Publication No. US20030219745A1  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Goodrich, Ryle  
APPLICANT: Liu, Chenghua  
APPLICANT: Ren, Feiyuan  
APPLICANT: Wang, Dunrui  
APPLICANT: Dmanac, Radoje T.  
TITLE OF INVENTION: No. US20030219745A1 Nucleic Acids and  
TITLE OF INVENTION: Polypeptides  
FILE REFERENCE: 802CON  
CURRENT APPLICATION NUMBER: US/10/120,988  
CURRENT FILING DATE: 2002-04-11

```
; PRIOR APPLICATION NUMBER: 09/774,528
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: pc_fl_genes Version 2.0
; SEQ ID NO 177
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (205)..(1305)
US-10-120-988-177

Alignment Scores:
Pred. No.: 1,47e-151      Length: 1416
Score: 1240.00           Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 17                   Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-120-988-177 (1-1416)

QY 1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLySerLeuAspIle 20
Db CTGTGGCATTAAGTTCAGAGAGATTTCAGATTTCCTGCTTGGAGAGTCTTGACATA 645
QY 21 AlaProGlyAspMetLeuAspLyValAlaArgArgLeuLeuIleLyHisProGlu 40
Db GACCCAGGTGACATGCTTGAAGAAGTGGCAAGAGACCTTCTTAAATAAATCCAGAG 705
QY 41 CysSerThrMetSerGlyLyValAlaIleGluHisLeuAlaLyGInGlyAsnArgPhe 60
Db TGTCTCCACCATGAGTGGGAGAAAGCCATAGACATTGGCCAAACAGAAATAGATT 765
QY 61 HisPheAspIleLySerProLeuHisHisAlaLyAsnCysAspPheSerPheThrGly 80
Db CATTTGACATCAAACTCCCTTGCATCATGCTAAATTTGATTTTCTTTACTGGA 825
QY 81 LeuGlnHisValThrAspLyIleIleMetLySgLySgLySgLyIleGluLyS 100
Db CTTCAACACCTTCTGATTAATAATATATGAAAAAGAGAGAGATTGAGAGAG 885
QY 101 GlnGlnIleLeuSerSerAlaIleAspIleAlaIleThrValGlnHisThrMetAlaCys 120
Db GGGCAAAATCTGCTTCAAGCAGACATGCTGCCACAGACAGCACATGCGCATGT 945
QY 121 HisLeuValLySArgThrHisArgAlaIleLeuPheCysLySgLySgLySgLySg 140
Db CACTTGTGAAAAAGACACATCGGCGCATTTCTGTTTGTAAAGCAGAGACTTGTACT 1005
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsnPheTyrlleArgArg 160
Db CAAAAATATATGACATGCTGTTGATCTGTGTGTCCAGATGACTTCTATATCCGAGA 1065
QY 161 AlaLeuGlnIleLeuThrAsnAlaIleThrGlnCysThrLeuLeuCysProProProArgLeu 180
Db GCTCTGAAATTTTAAACAAAGCAGACATGCTGTTGTGTCTCTCCCAAGACTA 1125
QY 181 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db TGCACGTATATATGACATTAATGATTCATGAGAAATGATTAAGAACTACGCTGGCTTG 1185
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrlleArgTyrlleArgTyrlleArgTyrlle 220
Db GGCATTTTATGATGACATAGAGAGCATCGCTATGAAACCAAAATGTCTCTTGGAGTAGAC 1245
QY 221 HisSerLySgLyValGlyValAserIleLySgLyValProGlnLeuLySmetGluIle 239
Db 1246 ATATCAAAAGAAAGTGGAGAGCTTCATATAAGTACACAAATTTAAAAATGAGAGATA 1302

RESULT 2
US-10-067-443-23
```

```
; Sequence 23, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEINASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-23

Alignment Scores:
Pred. No.: 1.64e-151      Length: 1526
Score: 1240.00           Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 14                   Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-23 (1-1526)

QY 1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLySerLeuAspIle 20
Db CTGTGGCATTAAGTTCAGAGAGATTTCAGATTTCCTGCTTGGAGAGTCTTGACATA 144
QY 21 AlaProGlyAspMetLeuAspLyValAlaArgArgLeuLeuIleLyHisProGlu 40
Db GACACAGGTGACATGCTTGAAGAAGTGGCAAGAGACCTTTTATATAAATCCAGAG 204
QY 41 CysSerThrMetSerGlyLyValAlaIleGluHisLeuAlaLyGInGlyAsnArgPhe 60
Db TGTCTCCACCATGAGTGGGAGAAAGCCATAGACATTGGCCAAACAGAAATAGATT 264
QY 61 HisPheAspIleLySerProLeuHisHisAlaLyAsnCysAspPheSerPheThrGly 80
Db CATTTGACATCAAACTCCCTTGCATCATGCTAAATTTGATTTTCTTTACTGGA 324
QY 81 LeuGlnHisValThrAspLyIleIleMetLySgLySgLySgLySgLyIleGluLyS 100
Db CTTCAACACCTTCTGATTAATAATATATGAAAAAGAGAGAGATTGAGAGAG 384
QY 101 GlnGlnIleLeuSerSerAlaIleAspIleAlaIleThrValGlnHisThrMetAlaCys 120
Db GGGCAAAATCTGCTTCAAGCAGACATGCTGCCACAGACAGCACATGCGCATGT 444
QY 121 HisLeuValLySArgThrHisArgAlaIleLeuPheCysLySgLySgLySgLySg 140
Db CACTTGTGAAAAAGACACATCGGCGCATTTCTGTTTGTAAAGCAGAGACTTGTACT 504
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsnPheTyrlleArgArg 160
Db CAAAAATATATGACATGCTGTTGATCTGTGTGTCCAGATGACTTCTATATCCGAGA 564
QY 161 AlaLeuGlnIleLeuThrAsnAlaIleThrGlnCysThrLeuLeuCysProProProArgLeu 180
Db GCTCTGAAATTTTAAACAAAGCAGACATGCTGTTGTGTCTCTCCCAAGACTA 624
QY 181 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db TGCACGTATATATGACATTAATGATTCATGAGAAATGATTAAGAACTACGCTGGCTTG 684
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrlleArgTyrlleArgTyrlleArgTyrlle 220
```

Db 685 GGCATTTTACATGACATGAGGACATCCGCTATGACCAAAATGCTCTTGGAGTAGAC 744  
Qy 221 IleserlysgluValGlyGluAlaSerIlelyValProGlnleuIleuysMetGluIle 239  
Db 745 AATATCAAAAGAAAGTTGGAGAGCTTCATTAAGTACCAATTAATAAATGAGATA 801

RESULT 3  
US-10-649-273-23  
Sequence 23, Application US/10649273  
Publication No. US20040043407A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
FILE REFERENCE: D0073 CNT  
CURRENT APPLICATION NUMBER: US/10/649,273  
CURRENT FILING DATE: 2003-08-27  
PRIOR APPLICATION NUMBER: US 60/266,518  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: US 10/067,443  
PRIOR FILING DATE: 2002-02-05  
PRIOR APPLICATION NUMBER: US 60/282,814  
PRIOR FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 23  
LENGTH: 1526  
TYPE: DNA  
ORGANISM: homo sapiens  
US-10-649-273-23

Alignment Scores:  
Pred. No.: 1 64e-151 Length: 1526  
Score: 1240.00 Matches: 239  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0  
DB: 18

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-649-273-23 (1-1526)

Qy 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLySerLeuAspIle 20  
Db 85 CTCTGGCATTAATTCAGAGAGTTTCAGATTTTCTGCTTCAGAAAGTCTTGGACATA 144  
Qy 21 AlaProGlyAspMetLeuAspLyValAlaArgArgLeuSerLeuIlelyshIspProGlu 40  
Db 145 GCACCAAGGTGACATGCTTGAAGAAGTGGCAAGACATTTCTTAATAAATCCAGAG 204  
Qy 41 CysSerThrMetSerGlyGlyValAlaIleGlnHisLeuAlaValysGlnGlyAsnArgPhe 60  
Db 205 TGCCTCCACCAATGAGTGTGGGAAAGCCATAGAACATTTGGCCAAGCAAGAAATAGATT 264  
Qy 61 HisPheAspIlelySProProLeuHisIleAlaIysAsnCysAspPheSerPheThrGly 80  
Db 265 CATTTGACATCAAACTCCCTTCATCATGCTAAATTTGTGATTTTCTTTTACTGGA 324  
Qy 81 LeuGlnHisValThrAspLyIleIleMetLySlyGlyLyGlnGlyIleGluIys 100  
Db 325 CTTCAACAGCTTAATTAATAATATGAAAGAAAGAAAGCAAGATTTGAGAG 384  
Qy 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120  
Db 385 GGGCAAAATCTGCTTCAGACAGACATGCTCCACAGTACAGACACAAATGCAATGT 444  
Qy 121 HisLeuValIysArgThrHisArgAlaIleLeuPheCysLySGlnArgAspLeuLeuPro 140  
Db 445 CATCTTGTAAGAAAGAACATCGGCTAATTCGTGTTGTAAGCAAGAGACTTGTACT 504  
Qy 141 GlnAsnAsnAlaValIleuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArg 160  
Db 505 CAAATTAATGCAAGTACTGCTTGCATCTGCTGCTGCTGCAATTAATTTATATCCGAGA 564  
Qy 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180

Db 565 GCTCTGAAATTTTAACAAAGCAACAGTGCCTTGTGTGCTCTCCACAGACTA 624  
Qy 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
Db 625 TGCACTGTAATGCACTTAATATATGATGATGATTAAGAAAGCTAGTGTGCTTG 684  
Qy 201 GlyIleLeuHisAspIleGlnGlyIleArgArgLeuProLySProLeuGlyValAsp 220  
Db 685 GGCATTTTACATGACATGAGAGCATCCGCTATGAAACCAAAATGCTCTTGGAGTAGAC 744  
Qy 221 IleserlysgluValGlyGluAlaSerIlelyValProGlnleuIleuysMetGluIle 239  
Db 745 AATATCAAAAGAAAGTTGGAGAGCTTCATTAAGTACCAATTAATAAATGAGATA 801

RESULT 4  
US-10-651-722-23  
Sequence 23, Application US/10651722  
Publication No. US20040048302A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
FILE REFERENCE: D0073 DIV  
CURRENT APPLICATION NUMBER: US/10/651,722  
CURRENT FILING DATE: 2003-08-29  
PRIOR APPLICATION NUMBER: US 60/266,518  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: US 10/067,443  
PRIOR FILING DATE: 2002-02-05  
PRIOR APPLICATION NUMBER: US 60/282,814  
PRIOR FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 23  
LENGTH: 1526  
TYPE: DNA  
ORGANISM: homo sapiens  
US-10-651-722-23

Alignment Scores:  
Pred. No.: 1 64e-151 Length: 1526  
Score: 1240.00 Matches: 239  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0  
DB: 18

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-651-722-23 (1-1526)

Qy 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLySerLeuAspIle 20  
Db 85 CTCTGGCATTAATTCAGAGAGTTTCAGATTTTCTGCTTCAGAAAGTCTTGGACATA 144  
Qy 21 AlaProGlyAspMetLeuAspLyValAlaArgArgLeuSerLeuIlelyshIspProGlu 40  
Db 145 GCACCAAGGTGACATGCTTGAAGAAGTGGCAAGACATTTCTTAATAAATCCAGAG 204  
Qy 41 CysSerThrMetSerGlyGlyValAlaIleGlnHisLeuAlaValysGlnGlyAsnArgPhe 60  
Db 205 TGCCTCCACCAATGAGTGTGGGAAAGCCATAGAACATTTGGCCAAGCAAGAAATAGATT 264  
Qy 61 HisPheAspIlelySProProLeuHisIleAlaIysAsnCysAspPheSerPheThrGly 80  
Db 265 CATTTGACATCAAACTCCCTTCATCATGCTAAATTTGTGATTTTCTTTTACTGGA 324  
Qy 81 LeuGlnHisValThrAspLyIleIleMetLySlyGlyLyGlnGlyIleGluIys 100  
Db 325 CTTCAACAGCTTAATTAATAATATGAAAGAAAGAAAGCAAGATTTGAGAG 384  
Qy 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120  
Db 385 GGGCAAAATCTGCTTCAGACAGACATGCTCCACAGTACAGACACAAATGCAATGT 444

|    |     |   |     |
|----|-----|---|-----|
| QY | 121 | HisLeuValIysArgThrHisArgAlaIleLeuPheCysIleGlnArgAspLeuPro                 | 140 |
| Db | 445 | CATCTTGGAAAGAAACACATCGGGCTATTCTGTTTGTGAGAGAGACCTGTTRACT                   | 504 |
| QY | 141 | GlnAsnAsnAlaValIleuValAlaSerGlyGlyValAlaSerAspPheTyrIleArgArg             | 160 |
| Db | 505 | CAAAATATATGACGACTGCTGGTTCATCTGGTGGTGTGCAAGTAACTTCTATATCCGAGA              | 564 |
| QY | 161 | AlaIleuGlnIleLeuThrAsnAlaThrGlnCysThrLeuIleuCysProProProArgLeu            | 180 |
| Db | 565 | GCTCTGGAATTTTAAACAACGCAACAACAGTGCACTTTGTTGTGTCCTCCCAAGCTA                 | 624 |
| QY | 181 | CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuAlaGlyLeu                 | 200 |
| Db | 625 | TGCATCTGATATATGGCATTAATGATTGTCATGGAATGTGATTAAGAAAGACTACCTGCTGGCTTG        | 684 |
| QY | 201 | GlyIleIleuHisAspIleGlnGlnGlyIleArgGlyArgIleProIlyseCysProIleuGlyValAlaAsp | 220 |
| Db | 685 | GGCATTATTAACATGACATATGAAAGGCAATCCGCTATGAACCAAAATGTCCTCTTGGAGTAGAC         | 744 |
| QY | 221 | IleSerIysGluValGlyValAsaSerIleIysValProGlnIleuIysMetGluIle                | 239 |
| Db | 745 | ATATCAAAAGAAAGTTGGAGAGACCTTCATTAAGAGTACCAACATTTAAATAATGGAGATA             | 801 |

```

RESULT 5
US-10-067-443-1
: Sequence 1, Application US/10067443
: Publication No. US20030082782A1
: GENERAL INFORMATION:
: APPLICANT: Bristol-Myers Squibb Company
: TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I
: TITLE OF INVENTION: SPINAL CORD, MP-1
: FILE REFERENCE: D0073 NP
: CURRENT FILING DATE: US/10/067,443
: PRIOR FILING DATE: 2002-02-05
: PRIOR APPLICATION NUMBER: US 60/266,518
: PRIOR FILING DATE: 2001-02-05
: PRIOR APPLICATION NUMBER: US 60/282,814
: PRIOR FILING DATE: 2001-04-10
: NUMBER OF SEQ ID NOS: 71
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 1
: LENGTH: 2197
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (231)..(1472)
: US-10-067-443-1

```

|   |           |
|---|-----------|
| Alignment Scores:   |           |
| Pred. No.:  | 2,79e-151 |
| Score:  | 1240.00   |
| Percent Similarity:   | 100.00%   |
| Best Local Similarity:  | 100.00%   |
| Query Match:  | 100.00%   |
| DB:   | 14        |
| US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-1 (1-2197) |           |
| Length:   | 2197      |
| Matches:  | 239       |
| Mismatched:   | 0         |
| Mismatches:   | 0         |
| Indels:   | 0         |
| Gaps:   | 0         |

|    |     |   |     |
|----|-----|---|-----|
| Qy | 1   | LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIle  | 20  |
| Db | 756 | CTGTGGCAATTAGTTCAAGAGATTTCAGATTTCTCTCTTTGGAAAGCTTTGGACATA     | 815 |
| Qy | 21  | AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysIleSproGlu | 40  |
| Db | 816 | GCACCAAGGTGACATGCTTGCACAAAGGTGGCAAGAGACTTTCTTAATATAACATCCAGAG | 875 |
| Qy | 41  | CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe  | 60  |
| Db | 876 | TGCTCCACCATTAAGTGGTGGAAAGCATATGGAACATTTGGCCAAACAAGAAATAGATTT  | 935 |
| Qy | 61  | HisPheAspIleLysProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly     | 80  |

Db 996 CATTTTGACATCAAAACCTCCCTGCATCATGCTAAATAATGTGATTTTCTTTTACTTCGGA 995

Qy 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIleGluLys 100

Db 996 CTTCAACACGTTACTGATTAATAATTAATGAAAAAGAAAAAGAGAGATATYGGAAG 105

Qy 101 GlyGlnIleLeuSerSerAlaIleAspIleAlaIleThrValGlnHisIleMetAlaCys 120

Db 1056 GGGCAAAATCCTGCTCTTCAGCGACAACATGCTGCACAGTACGACACACATATGGCATGT 111

Qy 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140

Db 1116 CATCTTGAGAAAGAACACATCGGGCTATTCGTTTGTAAACAGAGAGACTGTAACT 117

Qy 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160

Db 1176 CAAATATATGCAAGTACGTGGTGCATCTGGTGGTGCAGATGAACCTTATATATCCACAGA 123

Qy 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180

Db 1236 GCTCTGGAATTTTAACAAACGCAACACAGTCACTTGTGTGTCTCTCCACAGCTA 129

Qy 181 CysThrAspAsnGlyIleMetIleAlaTyrAsnGlyIleGluArgLeuArgAlaGlyLeu 200

Db 1236 TGCACTGATTAATGCGATTAATGATTCGATCGAATGGATTTGAAAGACTACGTGCGCTTG 135

Qy 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220

Db 1356 GGCATTTTATCATGATAGAGAGCGATCCGCTATGGAACCAAAATGTCTCTTGAGATGAGAC 141

Qy 221 ILeSerLysGluValGlyGlnAlaSerIleLysValProGlnLeuLysMetGlnIle 239

Db 1416 ATATCAAAAGAGTGTGAGAGACTTCCATTAAGAATGCCAATTTAAATAATGGAGATA 1472

```

RESULT 6
US-10-649-273-1
; Sequence 1, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
;

```

|                        |           |                 |
|------------------------|-----------|-----------------|
| Alignment Scores:      |           |                 |
| Pred. No.:             | 2,79e-151 | length: 2197    |
| Score:                 | 1240.00   | Matches: 239    |
| Percent Similarity:    | 100.00%   | Conservative: 0 |
| Best Local Similarity: | 100.00%   | Mismatches: 0   |
| Query Match:           | 100.00%   | Indels: 0       |
| DB:                    | 18        | Gaps: 0         |



```

Db      756 CTGTTGGCATTAGTTAGTCAAGAGATTTCAGATTTCCTCTTGGAAAGCTTTGGACATA 815
Qy      21  ALaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
Db      816 GCACCAAGGTGACATGCTTGACAGGTGGCAAGAACCTTTCTTTATATAACATCCAGAG 875
Qy      41  CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
Db      876 TGCTCCACCATGATGCTGGGAAAGCCATAGAACATTTGGCCAAACAAGAAATAGATTT 935
Qy      61  HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
Db      936 CATTGTGACATCAAACTCCCTTGACATCATGCTAAATAATGTGATTTTCTTTTACCTGA 995
Qy      81  LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100
Db      996 CTTCAACACGTTACTATATAAATATATATAAAGAAAGAAAGAAAGGATTTAGAGAG 1055
Qy      101  GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
Db      1056 GGGCAAAATCTGCTCTTACGACAGACATTTGCTCCACATACAGCACAAATGGCATGT 1115
Qy      121  HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140
Db      1116 CATCTTGTAAGAAAGAACACATCGGGCTATTCTGTTTGTAAAGCAGAGAGACTTGTACT 1175
Qy      141  GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160
Db      1176 CAAATAAATGCAAGTACGTGTTGATCGTGTGTCGCAAGTAACTTCTATATCCGAGA 1235
Qy      161  AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
Db      1236 GCTCTGAAATTTTAAACAACGACACATCGGCTATTCTGTTTGTAAAGCAGAGACTTGTACT 1295
Qy      181  CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db      1296 TGCACGTATATGCGATTATGATTGATGCAATGTAATGTAAGACATACGCTGGCTTG 1355
Qy      201  GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220
Db      1356 GGCATTTTACATGACATAGAGAGCATCCGCTATGAAACAAATGTCTCTTGAGAGTAGAC 1415
Qy      221  IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
Db      1416 ATATCAAAAGAAAGTTGGAAGAGCTTCATATAAAGTACCAATTAATAATGAGATA 1472

RESULT 7
US-10-651-722-1
; Sequence 1, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OR INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; PRIOR FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
US-10-651-722-1

```

```

Alignment Scores:
Pred. No.: 2,79e-151 Length: 2197
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 18 Gaps: 0

US-10-649-273-2_copy_176_414 (1-239) x US-10-651-722-1 (1-2197)

Qy      1  LeuLeuAlaLeuValGlnGlyValaSerAspPheLeuLeuGlyLysSerLeuAspIle 20
Db      756 CTGTTGGCATTAGTTAGTCAAGAGATTTCAGATTTCCTCTTGGAAAGCTTTGGACATA 815
Qy      21  ALaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
Db      816 GCACCAAGGTGACATGCTTGACAGGTGGCAAGAACCTTTCTTTATATAACATCCAGAG 875
Qy      41  CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
Db      876 TGCTCCACCATGATGCTGGGAAAGCCATAGAACATTTGGCCAAACAAGAAATAGATTT 935
Qy      61  HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
Db      936 CATTGTGACATCAAACTCCCTTGACATCATGCTAAATAATGTGATTTTCTTTTACCTGA 995
Qy      81  LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100
Db      996 CTTCAACACGTTACTATATAAATATATATAAAGAAAGAAAGAAAGGATTTAGAGAG 1055
Qy      101  GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
Db      1056 GGGCAAAATCTGCTCTTACGACAGACATTTGCTCCACATACAGCACAAATGGCATGT 1115
Qy      121  HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140
Db      1116 CATCTTGTAAGAAAGAACACATCGGGCTATTCTGTTTGTAAAGCAGAGAGACTTGTACT 1175
Qy      141  GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160
Db      1176 CAAATAAATGCAAGTACGTGTTGATCGTGTGTCGCAAGTAACTTCTATATCCGAGA 1235
Qy      161  AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
Db      1236 GCTCTGAAATTTTAAACAACGACACATCGGCTATTCTGTTTGTAAAGCAGAGACTTGTACT 1295
Qy      181  CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db      1296 TGCACGTATATGCGATTATGATTGATGCAATGTAATGTAAGACATACGCTGGCTTG 1355
Qy      201  GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220
Db      1356 GGCATTTTACATGACATAGAGAGCATCCGCTATGAAACAAATGTCTCTTGAGAGTAGAC 1415
Qy      221  IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
Db      1416 ATATCAAAAGAAAGTTGGAAGAGCTTCATATAAAGTACCAATTAATAATGAGATA 1472

RESULT 8
US-10-480-988-36
; Sequence 36, Application US/10480988
; Publication No. US20050069877A1
; GENERAL INFORMATION:
; APPLICANT: GANDHI, Ameena R.; KABIR, Amy B.;
; APPLICANT: SWARNAKAR, Anita; HAPALIA, April J.A.;
; APPLICANT: TRAN, Bao; DUGGAN, Brendan M.;
; APPLICANT: WARREN, Bridget A.; ISON, Craig H.;
; APPLICANT: HONCHRIT, Cynthia D.; MEYER, Daniel P.;
; APPLICANT: LU, Dying Aina M.; LEE, Ernestine A.;
; APPLICANT: YUE, Henry; FORSYTHE, Ian J.;
; APPLICANT: BARROSO, Ines; RAMKUMAR, Jayalaxmi;
; APPLICANT: GRIFPIN, Jennifer A.; LI, Joana X.;

```

APPLICANT: YANG, Junning; THANGAVELU, Kavitha;  
APPLICANT: GIETZEN, Kimberly J.; DING, Li;  
APPLICANT: BAUGHN, Mariah R.; BOROWSKY, Mark L.;  
APPLICANT: YAO, Monique G.; CHAMLA, Narinder K.;  
APPLICANT: MASON, Patricia M.; GURURAJAN, Rajagopal;  
APPLICANT: LEE, Sally; BECHA, Shanya D.;  
APPLICANT: LEE, Soo Yeun; TRAN, Uyen K.;  
APPLICANT: ELIOT, Vicki S.; LEO, Wen;  
APPLICANT: SPRAGUE, William W.; TANG, Y. Tom;  
APPLICANT: LU, Yan; ZEBARJADIAN, Yeganeh  
TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES  
FILE REFERENCE: PF-1040 USN  
CURRENT APPLICATION NUMBER: US/10/480, 988  
CURRENT FILING DATE: 2003-12-16  
PRIOR APPLICATION NUMBER: PCT/US02/19360  
PRIOR FILING DATE: 2002-06-18  
PRIOR APPLICATION NUMBER: US 60/300,508  
PRIOR FILING DATE: 2001-06-22  
PRIOR APPLICATION NUMBER: US 60/303,445  
PRIOR FILING DATE: 2001-07-06  
PRIOR APPLICATION NUMBER: US 60/305,405  
PRIOR FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: US 60/311,442  
PRIOR FILING DATE: 2001-08-09  
PRIOR APPLICATION NUMBER: US 60/314,821  
PRIOR FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 60/315,992  
PRIOR FILING DATE: 2001-08-29  
PRIOR APPLICATION NUMBER: US 60/378,205  
PRIOR FILING DATE: 2002-05-03  
NUMBER OF SEQ ID NOS: 56  
SOFTWARE: PERL Program  
SEQ ID NO 36  
LENGTH: 2572  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No: 7632424CB1  
US-10-480-988-36

Alignment Scores:  
Pred. No.: 3,52e-151 Length: 2572  
Score: 1240.00 Matches: 239  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 21 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-480-988-36 (1-2572)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
DB 669 CTGTTGGCATTTAGTTCAAGAGATTTCAGATTTCTGCTTGGAAAGCTTTGGACATA 728  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgLeuSerLeuIleLysHisProGlu 40  
DB 729 GCACCAAGTGACCTGCTTGAACAAGTGCGCAAGAACTTTCTTAATAAACAATCCAGAG 788  
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
DB 789 TCGTCCACCATGAGTGTGGGAAAGCCATAGAACATTTGGCCCAACAAGAAATAGATTT 848  
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80  
DB 849 CATTTTGCATCAAACTCCCTTCATCATGCTAATAAATGTGATTTTCTTTTATCGGA 908  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlyLysGlnGlyLysGlyLys 100  
DB 909 CTTCAACAGATTACTGATTAATAATAAATAAAGAAAGAAAGAGATATTGGAAG 968  
QY 101 GlyGlnIleLeuSerSerIleAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120

DB 969 GGGCAAACTCTGCTTGCAGACAGACATTCCTCCACAGTACAGACACAATGCGATGT 1028  
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
DB 1029 CATTTGTGAAAAGAACACATCGGGCTATTCTGTTTGTAAAGACAGAGACTTTGACT 1088  
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspPheTyrlleArgArg 160  
DB 1089 CAAATATATCAGATGATGCTGTGATGCTGTGCTCCAGATACTTATATCCGACAGA 1148  
QY 161 AlaGlnGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
DB 1149 GCTTGGAATTTTAAACAAGCAACAGGACATTTGTGTCTCTCCACAGACTA 1208  
QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
DB 1209 TGCATCTATATGCGATTAATGATTCATGAAATGTATGAAAGACATACGCTGCTTG 1268  
QY 201 GlyIleLeuHisAspIleGlyLysIleArgTyrGlnProLysCysProLeuGlyValAsp 220  
DB 1269 GGCATTTTACATGACATAGAAAGGATCGCTATGAAACCAAAATGTCTTGGAGTAGAC 1328  
QY 221 HisSerLysGlnValGlyGlnAlaSerIleLysValProGlnLeuLysMetGluIle 239  
DB 1329 ATATCAAAAGAGTTGAGAAAGCTTCATTAAGTACCAATTAATAAATGAGATA 1385

## RESULT 9

US-10-067-443-21  
Sequence 21, Application US/10067443  
Publication No. US20030082782A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN  
FILE REFERENCE: D0073 NP  
CURRENT APPLICATION NUMBER: US/10/067,443  
CURRENT FILING DATE: 2002-02-05  
PRIOR APPLICATION NUMBER: US 60/266,518  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: US 60/282,814  
PRIOR FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 21  
LENGTH: 1387  
TYPE: DNA  
ORGANISM: homo sapiens  
US-10-067-443-21

Alignment Scores:  
Pred. No.: 4,21e-148 Length: 1387  
Score: 1213.50 Matches: 238  
Percent Similarity: 90.15% Conservative: 0  
Best Local Similarity: 97.86% Mismatches: 1  
Query Match: 97.86% Indels: 25  
DB: 14 Gaps: 1

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-067-443-21 (1-1387)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
DB 549 CTGTTGGCATTTAGTTCAAGAGATTTCAGATTTCTGCTTGGAAAGCTTTGGACATA 608  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgLeuSerLeuIleLysHisProGlu 40  
DB 609 GCACCAAGTGACCTTGAACAAGTGCGCAAGAACTTTCTTAATAAACAATCCAGAG 668  
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
DB 669 TCGTCCACCATGAGTGTGGGAAAGCCATAGAACATTTGGCCCAACAAGAAATAGATTT 728  
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80

Db 729 CATTGGACATCAACCTCCCTGCATCATGCTAAATTTGATTTTCTTTACTGGA 788  
Qy 81 LeuGlnHisValThrAspLysIleIleMetLysGluysGluGluGlyIle----- 98  
Db 789 CTTCAACACCTTACTGATTAATATATATGAAAAAGAGAGAGATATATTTCTA 848  
Qy 98 ----- 98  
Db 849 ATTAGTAAAGTTGAACAGATAAATATTCCTGATTTGCTTAAATAAGCTCATTTTC 908  
Qy 99 -----GluysGluGlnIleLeuSerSerIleIleAspIleIleAlaThrValGln 115  
Db 909 TGCAGGTATGAGAGGGCAAAATCCTGCTTCAGACAGCATTTGCGCCACAGTACAG 968  
Qy 116 HisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysGln 135  
Db 969 CACACATGCGCATCTCATCTTGTGAAAAGAACACATCGGCTATTCGTTTGTGACAG 1028  
Qy 136 ArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsn 155  
Db 1029 AGAGACTGTTTACTCTCAAAATTAATGACGTACTGTTGCATCTGGTGTGCAAGTAC 1088  
Qy 156 PheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCys 175  
Db 1089 TTCTATATCCGACAGCTCTGGAATTTTAAACCAACACACAGTGCATCTTGTGT 1148  
Qy 176 ProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArg 195  
Db 1149 CTTCTCCCACTGACATGCTGATTAATGATGATGATGATGATGATGATGATGATG 1208  
Qy 196 LeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCys 215  
Db 1209 CTACGCTGCTGCTTGGCATTTTACATGACATAGAACATCCGCTATGAAACCAAAATGT 1268  
Qy 216 ProLeuGlyValAspIleSerLysGluValGlyLysAspIleLysValProGlnLeu 235  
Db 1269 CCTCTTGAGTAAACATATCAAAAGAGTGGAGAGCTTCATATAAAGTACCAATTA 1328  
Qy 236 LysMetGluIle 239  
Db 1329 AAAATGAGATA 1340  
RESULT 10  
US-10-649-273-21  
; Sequence 21, Application US/10649273  
; Publication No. US20040043407A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1  
; FILE REFERENCE: D0073 CNT  
; CURRENT APPLICATION NUMBER: US/10/649, 273  
; CURRENT FILING DATE: 2003-08-27  
; PRIOR APPLICATION NUMBER: US 60/266, 518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067, 443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282, 814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 21  
; LENGTH: 1387  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-649-273-21  
Alignment Scores:  
Pred. No.: 4,21e-148 Length: 1387  
Score: 1213.50 Matches: 238  
Percent Similarity: 50.15% Conservative: 0  
Best Local Similarity: 90.15% Mismatches: 1  
Query Match: 97.86% Indels: 25  
DB: 18 Gaps: 1

US-10-649-273-2\_copy\_176\_414 (1-239) x US-10-649-273-21 (1-1387)  
Qy 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
Db 549 CTTGTTGCGATTAAGTTCAAGAGAGTTTCAGATTTTTCGCTTCTTGAAAGCTTTGACATA 608  
Qy 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
Db 609 GCACACAGGTGACATGCTTGACAGAGTGGCAAGAGACCTTTCTTAATAAACATCCAGAG 668  
Qy 41 CysSerThrMetSerGlyGlyValAlaIleGlnHisLeuAlaLysGluGlyAsnArgPhe 60  
Db 669 TGCCTCACCATGAGTGTGGGAAAGCCATAGACATTTGGCCAAACAGAAATGATTT 728  
Qy 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
Db 729 CATTTGACATCAACCTCCCTTGATCATGCTAAATAATGTGATTTTCTTTTACTGGA 788  
Qy 81 LeuGlnHisValThrAspLysIleIleMetLysGluysGluGluGlyIle----- 98  
Db 789 CTTCAACACGTTACTGATTAATATATATGAAAAAGAGAGAGATATATTTCTA 848  
Qy 98 ----- 98  
Db 849 ATTAGTAAAGTTGAACAGATAAATATTCCTGATTTGCTTAAATAAGCTGCTCATTTTC 908  
Qy 99 -----GluysGluGlnIleLeuSerSerIleIleAspIleIleAlaThrValGln 115  
Db 909 TGCAGGTATGAGAGGGCAAAATCCTGCTTCAGACAGCATTTGCGCCACAGTACAG 968  
Qy 116 HisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysGln 135  
Db 969 CACACATGCGCATCTCATCTTGTGAAAAGAACACATCGGCTATTCGTTTGTGACAG 1028  
Qy 136 ArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsn 155  
Db 1029 AGAGCTTGTTACTCTCAAAATTAATGACGTACTGTTGCATCTGGTGTGCAAGTAC 1088  
Qy 156 PheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCys 175  
Db 1089 TTCTATATCCGACAGCTCTGGAATTTTAAACCAACACACAGTGCATCTTGTGT 1148  
Qy 176 ProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArg 195  
Db 1149 CTTCTCCCACTGACATGCTGATTAATGATGATGATGATGATGATGATGATGATG 1208  
Qy 196 LeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCys 215  
Db 1209 CTACGCTGCTGCTTGGCATTTTACATGACATAGAACATCCGCTATGAAACCAAAATGT 1268  
Qy 216 ProLeuGlyValAspIleSerLysGluValGlyLysAspIleLysValProGlnLeu 235  
Db 1269 CCTCTTGAGTAAACATATCAAAAGAGTGGAGAGCTTCATATAAAGTACCAATTA 1328  
Qy 236 LysMetGluIle 239  
Db 1329 AAAATGAGATA 1340  
RESULT 11  
US-10-651-722-21  
; Sequence 21, Application US/10651722  
; Publication No. US20040048302A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1  
; FILE REFERENCE: D0073 DIV  
; CURRENT APPLICATION NUMBER: US/10/651, 722  
; CURRENT FILING DATE: 2003-08-29  
; PRIOR APPLICATION NUMBER: US 60/266, 518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067, 443  
; PRIOR FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/282,814  
PRIOR FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: Patentin version 3.2  
SEQ ID NO 21  
LENGTH: 1387  
TYPE: DNA  
ORGANISM: homo sapiens  
US-10-651-722-21

## Alignment Scores:

| Pred. No.:             | 4,21e-148 | Length:       | 1387 |
|------------------------|-----------|---------------|------|
| Score:                 | 1213.50   | Matches:      | 238  |
| Percent Similarity:    | 90.15%    | Conservative: | 0    |
| Best Local Similarity: | 90.15%    | Mismatches:   | 1    |
| Query Match:           | 97.86%    | Indels:       | 25   |
| DB:                    | 18        | Gaps:         | 1    |

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-651-722-21 (1-1387)

1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
549 CTCTTGCACTACCTTCAAGAGATTTCGATTTCTGCTTGGAAAGCTTTGACATA 608  
21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
609 GCACCGAGTGACATGCTTGACAGAGTGCAAGAACTTTCTTAATTAACAATCCAGAG 668  
41 CysSerTherMetSerGlyLysValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
669 TGCCTCCACATGAGTGTGTGGAAAGCCATGAGCATTTGGCCAAACAGAAATAGATT 728  
61 HisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80  
729 CATTTTGACATCAACCTCCCTTGACATGCTTAATAATGATGATTTCTTTACTGGA 788  
81 LeuGlnHisValThrAspLysIleIleMetLysGlyLysGlnGlyLysIle----- 98  
789 CTTCAACAGCTTATGATTAATATATGAAAAAGAAAAAGAGAGATATATTTCTA 848  
98 ----- 98  
849 ATTAGTAAGTTGACAGATTAATATTTCTGATGTTGCTTAATAATAGCTGCTCATTTTC 908  
99 -----GluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGln 115  
909 TGCAGGATATGAGAGGGGCAATCTGCTTCCAGCAGACATGCTGCACAGTACAG 968  
116 HisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGln 135  
969 CACACATATGAGCATGCTATCTTGGAAGAAGACATCGGGCTATTTCTGTTTGTAAACAG 1028  
136 ArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyLysValAlaSerAsn 155  
1029 AGAGACTTGTATCTCAAAATATATGAGTCTGTTGATCTGTGTGTCGCAAGTAAAC 1088  
156 PheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCys 175  
1089 TTCTATATTCGCGACGCTCTGGAATTTTAACAACGCAACACAGTCCACTTTGTTGTGT 1148  
176 ProProArgArgLeuCysThrAspAsnGlyIleMetIleAlaThrPaenglyIleGluArg 195  
1149 CTTCTCCCAAGCATATGCACTGATATATGCAATTAATGATTCATGAGTAATGATTAAGA 1208  
196 LeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCys 215  
1209 CTACGAGTGTGCTTGGGCAATTTATCATGATGAGAGGCAATCCGCTATGAAACCAAAATGT 1268  
216 ProLeuGlyValAspIleSerLysGlyValGlyGluAlaSerIleLysValProGlnLeu 235  
1269 CTTCTTGAGTATGACATATCAAAAAGAAAGTGGAGAGCTTCCATTAATAAGTACCAATTA 1328  
236 LysMetGluIle 239

Db 1329 AAAATGAGATA 1340

## RESULT 12

US-10-887-553A-1047  
Sequence 1047, Application US/10887553A  
Publication No. US20050085436A1  
GENERAL INFORMATION:  
APPLICANT: Garza, Dan  
APPLICANT: Li, Hao

TITLE OF INVENTION: Method to treat conditions associated  
with insulin signalling dysregulation  
FILE REFERENCE: 4-33262  
CURRENT APPLICATION NUMBER: US/10/887,553A

CURRENT FILING DATE: 2004-07-08

PRIOR APPLICATION NUMBER: 60/485,883

PRIOR FILING DATE: 2003-08-07

NUMBER OF SEQ ID NOS: 1208

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1047

LENGTH: 1387

TYPE: DNA

ORGANISM: human

US-10-887-553A-1047

## Alignment Scores:

| Pred. No.:             | 4,21e-148                                | Length:       | 1387 |
|------------------------|--|---------------|------|
| Score:                 | 1213.50 <td>Matches:</td> <td>238</td>   | Matches:      | 238  |
| Percent Similarity:    | 90.15% <td>Conservative:</td> <td>0</td> | Conservative: | 0    |
| Best Local Similarity: | 90.15% <td>Mismatches:</td> <td>1</td>   | Mismatches:   | 1    |
| Query Match:           | 97.86% <td>Indels:</td> <td>25</td>      | Indels:       | 25   |
| DB:                    | 21                                       | Gaps:         | 1    |

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-887-553A-1047 (1-1387)

1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
549 CTCTTGCACTACCTTCAAGAGATTTCGATTTCTGCTTGGAAAGCTTTGACATA 608  
21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
609 GCACCGAGTGACATGCTTGACAGAGTGCAAGAACTTTCTTAATTAACAATCCAGAG 668  
41 CysSerTherMetSerGlyLysValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
669 TGCCTCCACATGAGTGTGTGGAAAGCCATGAGCATTTGGCCAAACAGAAATAGATT 728  
61 HisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80  
729 CATTTTGACATCAACCTCCCTTGACATGCTTAATAATGATGATTTCTTTACTGGA 788  
81 LeuGlnHisValThrAspLysIleIleMetLysGlyLysGlnGlyLysIle----- 98  
789 CTTCAACAGCTTATGATTAATATATGAAAAAGAAAAAGAGAGATATATTTCTA 848  
98 ----- 98  
849 ATTAGTAAGTTGACAGATTAATATTTCTGATGTTGCTTAATAATAGCTGCTCATTTTC 908  
99 -----GluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGln 115  
909 TGCAGGATATGAGAGGGGCAATCTGCTTCCAGCAGACATGCTGCACAGTACAG 968  
116 HisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGln 135  
969 CACACATATGAGCATGCTATCTTGGAAGAAGACATCGGGCTATTTCTGTTTGTAAACAG 1028  
136 ArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyLysValAlaSerAsn 155  
1029 AGAGACTTGTATCTCAAAATATATGAGTCTGTTGATCTGTGTGTCGCAAGTAAAC 1088  
156 PheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCys 175

DB 1089 TTCTATATCCGAGAGCTGTGAAATTTTAAACCAACACAGTCGACTTGTGTGT 1148  
QY 176 PROPROBARGLEUCYSTRASPANGIYILEMETLEALATPANGIYILEGLUARG 195  
DB 1149 CCCCCCAGAGATATGACATGATTAATGACATTAATGATTAATGATTAATGATTAATG 1208  
QY 196 LEUARGALAGIYLEUGIYILEUHIASPIILEGLIYILEARGIYRGUPLPOLYCY 215  
DB 1209 CTACGTGTGTGCTTGGGCACTTTTACATGACATGACATGACATGACATGACATG 1268  
QY 216 PROLEUGIYVALASPILLESERLYGIVUAGIYGLUVALASERILEYVALPROGLU 235  
DB 1269 CCTCTTGAGGTAGACATATATCAAAAAGAGTTGAGAACCTTCATTAACCAATTA 1328  
QY 236 LYSMETGLUTILE 239  
DB 1329 AATATGAGATA 1340  
RESULT 13  
US-10-012-140-6  
; Sequence 6, Application US/10012140  
; Publication No. US2003009017A1  
; GENERAL INFORMATION:  
; APPLICANT: Leiby, Kevin R.  
; APPLICANT: Kapeller-Libermann, Rosana  
; APPLICANT: Glucksmann, Maria A.  
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND  
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES  
; FILE REFERENCE: 381552004900  
; CURRENT APPLICATION NUMBER: US/10/012,140  
; PRIOR FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: 60/246,768  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/246,772  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/249,185  
; PRIOR FILING DATE: 2000-11-15  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 1245  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-012-140-6  
Alignment Scores:  
Pred. No.: 8,53e-147 Length: 1245  
Score: 1203.00 Matches: 232  
Percent Similarity: 98.33% Conservative: 3  
Best Local Similarity: 97.07% Mismatches: 4  
Query Match: 97.02% Indels: 0  
DB: 14 Gaps: 0  
US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-012-140-6 (1-1245)  
QY 1 LEUENALALEUVALGINGIYVALSERASPHELEULENGIYLYSESERLEUASPILE 20  
DB 526 CTGTGGCATTAAGTTCAGAGGTTTCAGATTTTCGCTTCTTGAGAACTTTTGGACATA 585  
QY 21 ALAPROGLIYASPHETLEUASPIYVALALARGARGLEUSERLEULEYHISPROGLU 40  
DB 586 GCACCGAGTACATGCTTGACACAGGTGCAAGAACCTTTTAAATMAAACATCCAGAG 645  
QY 41 CYSESETHMETSERGLYGLYVALAILEGLIHIISLEUALYVSGINGIYASARGPHE 60  
DB 646 TGTCTCCACCAAGTGGGGAGGAGCCATAGAACATTTGGCCAAACAGAAATAGATT 705  
QY 61 HISPEASPIILEYSPROPROLEUHIHISALALYSAENCYASAPHESESPHETHRGLY 80  
DB 706 CATTTTACATCAACCTCCCTTGCAATCACTAAATTTGATTTTCTTTTACTGGA 765  
QY 81 LEUGIHISVALTHRAAPLYLLEILEMETLYSGIULYSGIUGIYILEGLULYS 100

DB 766 CTTCAACCGCTTACTGATTAATAATGAAAACGGAACACAGAGAGTATTCAGAG 825  
QY 101 GIVGILHILEUSERSERALALASPIILEALALATHRVALGHIHETHMEALACY 120  
DB 826 GGGCAATTCCTGCTTCCAGCAGACATTCCTGCAAGATACACACATGSCATGT 885  
QY 121 HISLEUVALYARGTHRHISARGALALILEUPHECYSLYSGINARGASPLEUENPRO 140  
DB 886 CATCTTGAGAAAGAACACATCGGGCTATTCGTTTGTAAAGCAGAGACTTGTACCT 945  
QY 141 GINARSENALAVALLEUVALIASERGLYGLYVALALASERASPHEPTYILEARGRG 160  
DB 946 CAATAATATGACATGCTGTGACATCTGTGTGCGAAGTACTTATATTCGACGA 1005  
QY 161 ALALEUGIUILELEUTHRASNALATHRGINCYSTRHLEULEUCYSPROPROBARGLE 180  
DB 1006 GCTCTGAAATTTTAAACCAACGACACACAGTCGACTTGTGTGTCTCTCCAGACTA 1065  
QY 181 CYSTRASPANGIYILEMETLEALATPANGIYILEGLUARGLEUARGALAGIYLEU 200  
DB 1066 TGCACTGATATGACATTAATGATGATGATGATGATGATGATGATGATGATGATG 1125  
QY 201 GIVHILEUHIASPIILEGLIYILEARGIYRGUPLPOLYCYSPROLEUGIYVALASP 220  
DB 1126 GGCATTTTACATGACATGAGAGGATCCGCTATGAAACCAAAATGCTCTTGAGTAC 1185  
QY 221 ILESERLYSGIULYGLUVALASERILEYVALPROGLINLEUYSMETGLUTILE 239  
DB 1186 AATATCAAAAAGAGTTGAGAGGCTTCATTAATAAGTACCAATTAATAATGAGATA 1242  
RESULT 14  
US-10-012-140-4  
; Sequence 4, Application US/10012140  
; Publication No. US2003009017A1  
; GENERAL INFORMATION:  
; APPLICANT: Leiby, Kevin R.  
; APPLICANT: Kapeller-Libermann, Rosana  
; APPLICANT: Glucksmann, Maria A.  
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND  
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES  
; FILE REFERENCE: 381552004900  
; CURRENT APPLICATION NUMBER: US/10/012,140  
; PRIOR FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: 60/246,768  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/246,772  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/249,185  
; PRIOR FILING DATE: 2000-11-15  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 1820  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (146)...(1390)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1820)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-012-140-4  
Alignment Scores:  
Pred. No.: 1,49e-146 Length: 1820  
Score: 1203.00 Matches: 232  
Percent Similarity: 98.33% Conservative: 3  
Best Local Similarity: 97.07% Mismatches: 4  
Query Match: 97.02% Indels: 0  
DB: 14 Gaps: 0

```
US-10-649-273-2_COPY_176_414 (1-239) x US-10-012-140-4 (1-1820)
OY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB 671 CTTGTGGCATTATCTCAAGAGAGTTTGCATTTCTGCTTTGGAAGTCTTTGACATA 730
OY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 731 GCAACGAGTGACATGCTTGCAAGAGTGCGCAAGAGACTTTCTTAATAAACAATCCAGAG 790
OY 41 CysSerThrMetSerGlyGlyLysValAlaLeuHisLeuAlaLysGlnGlyAspArgPhe 60
DB 791 TGTCTCCACCAATGAGTGTGGGAAAGCCATAGAACATTGTGCGCAACAAGAAATAGATT 850
OY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80
DB 851 CATTGTGACATCAACCTCCCTTGCAATGCTAAATGCTGATTTTCTTTTACTGGA 910
OY 81 LeuGlnHisValThrAspLysIleIleMetLysGlyLysGlnGlyIleGluLys 100
DB 911 CTTCAACAGCTTACTGATATAAATAATGAAACAGAAACAGAAAGAGATGAGAG 970
OY 101 GlyGlnIleLeuSerSerIleAlaAspIleAlaAlaThrValGlnHisSerMetAlaCys 120
DB 971 GGGCAAAATCCTGCTTCAGCAGACGACATTGCTGCCACAGTACAGCACAAATGCGATGT 1030
OY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB 1031 CATCTGTGAAAAAGAACACATCGGGCTATTCTGTTTGTAAAGCAGAGACTTTGTTACT 1090
OY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160
DB 1091 CAAATAATAGCAGTACTGTTGTCATCGTGTGTGTGCAAGTAACTTCTATATCCGAGA 1150
OY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuGlyProProProArgLeu 180
DB 1151 GCTCTGGAATTTTAAACAACGCAACACAGTGCATTTGTTGTCTCTCCACAGACTA 1210
OY 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGlnArgLeuArgAlaGlyLeu 200
DB 1211 TGCACGATTAATGCGCATTAATGATTCATGGAATGTAATGTAAGACTACAGCTGGCTTG 1270
OY 201 GlyIleLeuHisAspIleGlnGlyIleArgTyrGlnProLysCysProLeuGlyValAsp 220
DB 1271 GGCATTTTACATGACATAGAGGCAATCCGCTATGAAACCAAAATGCTCTCTTGAGTGA 1330
OY 221 HisSerLysGlnValGlyGlnAlaSerIleLysValProGlnLeuLysMetGluIle 239
DB 1331 ATATCAAAAGAGAGTTGAGAGAGCTTCATTAAGTACCAATTAATAATGAGAGATA 1387

RESULT 15
US-10-094-749-400
; Sequence 400, Application US/10094749
; Publication No: US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, KYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHITO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
```

```
APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 400
; LENGTH: 2208
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-400

Alignment Scores:
Pred. No.: 1,42e-127 Length: 2208
Score: 1059.00 Matches: 211
Percent Similarity: 88.28% Conservative: 0
Best Local Similarity: 88.28% Mismatches: 4
Query Match: 85.40% Indels: 24
DB: 17 Gaps: 1

US-10-649-273-2_COPY_176_414 (1-239) x US-10-094-749-400 (1-2208)
OY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB 869 CTTGTGGCATTATCTCAAGAGAGTTTGCATTTCTGCTTTGGAAGTCTTTGACATA 928
OY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 929 GCAACAGTGACATGCTTGAACAAGTGGCAAGAAAGACTTCTTAATAAATCATCAGAG 988
OY 41 CysSerThrMetSerGlyGlyLysValAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 989 TGTCCACACATGAGTGTGGGAAAGCCATAGAACATTGTGCGCAACAAGAAATGATTT 1048
OY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80
DB 1049 CATTGTGACATCAAACTCCCTTGCAATGCTAAATATGTAATTTCTTTTACTGGA 1108
OY 81 LeuGlnHisValThrAspLysIleIleMetLysGlyLysGlnGlyIleGluLys 100
DB 1109 CTTCAACACGTTACTGATTAATAATGAAAGAAAGAGAGCTTTTTCAGAG 1168
OY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisSerMetAlaCys 120
DB 1169 GGGCAAAATCCTGCTTTCAGCAGACAGACTTCTCCACAGTACAGACAAATGGCATGT 1228
OY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB 1229 CATCTTGTGAAMAAACACATCGGGCTATTCTGTTTGTAAAGCAGAGACTTGTACT 1288
OY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160
DB 1289 CAAATAATGCAAGTCTGTTGCAATCGTGTGTGTGCGAAGTAACTTCTGATCCGACGA 1348
OY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuGlyProProProArgLeu 180
DB 1349 GCTCTGGAATTTTAAACAACAGCAACAGTGTGTTGTGTCTCTCCACAGACTA 1408
OY 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGlnArgLeuArgAlaGlyLeu 200
DB 1409 TGCACATTAATGCAATTAATGATTTGCA----- 1435
OY 201 GlyIleLeuHisAspIleGlnGlyIleArgTyrGlnProLysCysProLeuGlyValAsp 220
DB 1436 -----TCATGCTCTTGTGAGTAC 1456
OY 221 HisSerLysGlnValGlyGlnAlaSerIleLysValProGlnLeuLysMetGluIle 239
```

Db 1457 ATATCAAGAAGTTGGAGAAGCTTCCATTAAGTACCACATTTAAAAATGGAGATA 1513

Search completed: June 17, 2005, 08:03:30  
Job time : 1634.64 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Comugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: June 16, 2005, 18:30:49 ; Search time 205.65 Seconds  
(without alignments)  
3294.036 Million cell updates/sec

Title: US-10-649-273-2  
Perfect score: 2125  
Sequence: 1 MILITKAGVFPKSRKRVY.....DISKEVGEASIKVQLMEI 414

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 segs, 818138359 residues  
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seg length: 0  
Maximum DB seg length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ .p2n.model -DEV=x1p  
-Q=/cgn2\_1/USPTO.spool\_p/US10649273/runat\_15062005\_111417\_6057/app\_query.fasta\_1.1429  
-DB=Issued Patents NA -OPMT=fasta -SUFFIX=p2n.rnt -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blonsum62 -TRANS=human40.cdi  
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10649273.@CGN\_1.1.177 @runat\_15062005\_111417\_6057 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:  
5: /cgn2\_6/ptodata/1/ina/PTUS.COMB.seq:  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score  | Query Match | Length DB | ID                    | Description       |
|------------|--------|-------------|-----------|-----------------------|-------------------|
| 1          | 2125   | 100.0       | 2197      | 4 US-10-067-443-1     | Sequence 1, Appli |
| 2          | 2090.5 | 98.4        | 1387      | 4 US-10-067-443-21    | Sequence 21, Appl |
| 3          | 1747   | 82.2        | 1416      | 4 US-09-774-528-177   | Sequence 177, App |
| 4          | 1385   | 65.2        | 1526      | 4 US-10-067-443-23    | Sequence 23, Appl |
| 5          | 1186.5 | 55.8        | 14364     | 4 US-10-067-443-20    | Sequence 20, Appl |
| 6          | 502    | 23.6        | 94750     | 4 US-09-536-002-38    | Sequence 38, Appl |
| 7          | 501    | 23.6        | 1053      | 4 US-09-540-236-806   | Sequence 806, App |
| 8          | 499.5  | 23.5        | 1206      | 4 US-09-252-991A-801  | Sequence 801, App |
| 9          | 452.5  | 23.2        | 1059      | 4 US-09-252-991A-884  | Sequence 884, App |
| 10         | 482.5  | 22.7        | 1074      | 4 US-09-543-681A-2341 | Sequence 2341, Ap |
| 11         | 482    | 22.7        | 996       | 4 US-09-902-540-6612  | Sequence 6612, Ap |
| 12         | 482    | 22.7        | 2582      | 4 US-09-902-540-503   | Sequence 503, App |

|    |       |      |         |                       |                    |
|----|-------|------|---------|-----------------------|--------------------|
| 13 | 479   | 22.5 | 1315    | 1 US-08-087-797-1     | Sequence 1, Appli  |
| 14 | 475.5 | 22.4 | 1830121 | 4 US-09-557-884-1     | Sequence 1, Appli  |
| 15 | 475.5 | 22.4 | 1830121 | 4 US-09-643-990A-1    | Sequence 1, Appli  |
| 16 | 474.5 | 22.3 | 1032    | 4 US-09-489-039A-2050 | Sequence 2050, Ap  |
| 17 | 472   | 22.2 | 1092    | 4 US-09-107-532A-2955 | Sequence 2955, Ap  |
| 18 | 461   | 21.7 | 1008    | 3 US-08-987-121A-5    | Sequence 5, Appli  |
| 19 | 461   | 21.7 | 1011    | 3 US-08-987-121A-3    | Sequence 3, Appli  |
| 20 | 457   | 21.5 | 1006    | 3 US-08-961-083-51    | Sequence 51, Appl  |
| 21 | 457   | 21.5 | 1006    | 4 US-09-536-784-51    | Sequence 51, Appl  |
| 22 | 457   | 21.5 | 10974   | 3 US-08-961-527-214   | Sequence 214, App  |
| 23 | 455   | 21.4 | 1011    | 3 US-09-066-512-1     | Sequence 1, Appli  |
| 24 | 454   | 21.4 | 1011    | 4 US-09-134-000C-1551 | Sequence 1551, Ap  |
| 25 | 453   | 21.3 | 1011    | 4 US-09-583-110-2196  | Sequence 2196, Ap  |
| 26 | 448   | 21.1 | 1011    | 4 US-09-107-433-1618  | Sequence 1618, Ap  |
| 27 | 443   | 20.8 | 1107    | 3 US-09-134-001C-1072 | Sequence 1072, Ap  |
| 28 | 433   | 20.4 | 1026    | 3 US-09-149-624-1     | Sequence 1, Appli  |
| 29 | 424   | 20.0 | 15249   | 4 US-08-956-171E-102  | Sequence 102, App  |
| 30 | 424   | 20.0 | 15249   | 4 US-08-781-986A-102  | Sequence 102, App  |
| 31 | 406.5 | 19.1 | 640681  | 4 US-09-790-988-1     | Sequence 1, Appli  |
| 32 | 406.5 | 19.1 | 1230025 | 4 US-09-198-452A-1    | Sequence 1, Appli  |
| 33 | 406.5 | 19.1 | 1230230 | 4 US-09-438-185A-1    | Sequence 1, Appli  |
| 34 | 404.5 | 19.0 | 1155    | 4 US-09-602-777A-147  | Sequence 147, App  |
| 35 | 404   | 19.0 | 3064    | 3 US-09-221-017B-794  | Sequence 794, App  |
| 36 | 379.5 | 17.9 | 4403765 | 3 US-09-103-840A-2    | Sequence 2, Appli  |
| 37 | 379.5 | 17.9 | 4411529 | 3 US-09-103-840A-1    | Sequence 19, Appl  |
| 38 | 369.5 | 17.4 | 645     | 4 US-09-328-352-483   | Sequence 483, App  |
| 39 | 362   | 17.0 | 876     | 4 US-09-724-623-19    | Sequence 19, Appl  |
| 40 | 357   | 16.8 | 580073  | 4 US-08-545-528D-1    | Sequence 1, Appli  |
| 41 | 333.5 | 15.7 | 42325   | 4 US-08-311-721A-131  | Sequence 131, App  |
| 42 | 327.5 | 15.4 | 3993    | 4 US-09-710-279-3985  | Sequence 3985, App |
| 43 | 319   | 15.0 | 822     | 4 US-09-710-279-727   | Sequence 727, App  |
| 44 | 317   | 14.9 | 36941   | 4 US-08-311-731A-130  | Sequence 130, App  |
| 45 | 313   | 14.7 | 1664976 | 4 US-08-916-421B-1    | Sequence 1, Appli  |

## ALIGNMENTS

RESULT 1  
US-10-067-443-1  
; Sequence 1, Application US/10067443  
; Patent No. 6642041  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
; TITLE OF INVENTION: SPINAL CORD, ME-1  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 2197  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (231)..(1472)  
; US-10-067-443-1

Alignment Scores:  
Pred. No.: 5.72e-261  
Score: 2125.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 100.00%  
DB: 4  
US-10-649-273-2 (1-414) x US-10-067-443-1 (1-2197)

|               |      |
|---------------|------|
| Length:       | 2197 |
| Matches:      | 414  |
| Conservative: | 0    |
| Mismatches:   | 0    |
| Indels:       | 0    |
| Gaps:         | 0    |



QY 1 MetLeuIleLeuThrIleuThraAlaGlyValPhePheIysProSerIysaArgLysValTyr 20  
DB ATCTTAATCTTGAAGTAACTGACGAGAGTTTCTTTTAAACCATCAAAAGAAAGTTTAT 290  
QY 21 GluPheLeuArgSerPheLeuPheIleProGlyThrIleuPheLeuIleValIleu 40  
DB GAATTTTAAAGAAAGTTTAAATTTTCAATCCGGAACACTATTTCTTCAAAAATGTATG 350  
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaIleValIleAspGluThrGlyAsnVal 60  
DB GGAAATGAAACTGATTTGATGATGATGACGACCTCTGTGTGATGAACTCGAAATGTG 410  
QY 61 LeuGlyValIleAlaIleHisSerGlnThrGluValHisLeuIysThrGlyIleValPro 80  
DB TTGGAGAAACCAATACATTCCTCCAAAGTGAAGTTTAAAAAACAGGGGATTTCTCT 470  
QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
DB CCGAGAGCTCAACAGCTTCAAGAGAAATTCACAGAAATGATACAGAGAGCTCTTCT 530  
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrTrlIleYsProGlyLeu 120  
DB GCCAGTGAAGTCTCTCCAGTGAAGCTCTCAAGCAATTCACATCAATCAAAACAGAGCTT 590  
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuIleLys 140  
DB GCTTTAAGCTCGGAGAGGCTTATCATTTAGCTTACAGCTGTAGAGACAGTTAAAAAG 650  
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
DB CCATTCATTCCTCATTCATCATGAGAGCTCATGCACTTCACTTATGATTCAGCAATAAA 710  
QY 161 ValGluPheProPheLeuValIleLeuIleSerGlyIleHisCysLeuLeuAlaLeuVal 180  
DB TATGAAATTCCTTTTATGATTTTATGATTTTCTGAGAGTCACTGTGTGTGCAATGAT 770  
QY 181 GlnGlyValSerAspPheLeuLeuGlyIysSerLeuAspIleAlaProGlyAspMet 200  
DB CAAGAGTTTCAATTTTCTGCTTCTTGAAGCTTTTGAAGATACAGACAGGAGCATG 830  
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleIysHisProGlyCysSerThrMetSer 220  
DB CTTGACAGAGTGGCAAGAGACTTTCTTTAATAAATCATCCAGAGTGTCTCCACCATGAT 890  
QY 221 GlyGlyValAlaIleGlnIleLeuAlaLysGlnIysArgPheHisPheAspIleLys 240  
DB GGTGGGAAACCATGAGCAATTTGGCCAAACAGAAATGATTTCTTTGACATCAAA 950  
QY 241 ProProLeuHisIleAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
DB CTCTCCCTTGCAATGCTTAATAAATTTGATTTTCTTTTACTGACCTTCAACAGCTTACT 1010  
QY 261 AspLysIleIleMetLysLysGlyLysGlyGluGlyIleGlyLysGlyGlnIleLeuSer 280  
DB GATTAATAATAATAAGAAAGAAAGAAAGAGGATTTGAGAGGGGCAATCTGTCT 1070  
QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300  
DB TCAGCAGCAGACATGCTGTCACAGTACACACACCAATGAGATCTCTGTGAAAAAGA 1130  
QY 301 ThrHisArgAlaAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320  
DB ACACATGAGGCTATTTCTTTTGTAGACAGAGAACTTTGATCTCAAAATATGACGATA 1190  
QY 321 LeuValAlaSerGlyValAlaLysAspPheThrIleArgArgAlaLeuGlnIleLeu 340  
DB CTGTGTGCAATCTGTGTGTGTGTGCAAGTACTTCAATTCGACAGAGCTCTGGAATTTTA 1250  
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaArgLeuCysThrAspAsnGly 360  
DB ACAAGCAGCAACAGTGCATTTGTGTGTGTCTCTCCAGACATATGACATGATTAATGCC 1310  
QY 361 IleMetIleAlaIleArgLysGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380

DB 1311 ATTAATGATTCATGCAATGATGATTAAGAAAGTACAGTGTGCTTGGCATTTTACATGAC 1370  
QY 381 IleGluGlyIleArgGlyArgLysCysProLeuGlyValAspIleSerLysGluVal 400  
DB ATAGAGGCAATCCCTATGAAACCAAAATGCTCTTGTGAGTACATATACAAAGAAAGTT 1430  
QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414  
DB GGAAAGCTTCATTAATAAGTACCAATTTAAATAATGAGATA 1472  
RESULT 2  
US-10-667-443-21  
Sequence 21, Application US/10067443  
Patent No. 6642041  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN  
FILE REFERENCE: D0073 NP  
CURRENT APPLICATION NUMBER: US/10/067,443  
PRIOR FILING DATE: 2002-02-05  
PRIOR APPLICATION NUMBER: US 60/266,518  
PRIOR FILING DATE: 2001-02-05  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 21  
LENGTH: 1387  
TYPE: DNA  
ORGANISM: homo sapiens  
US-10-667-443-21  
Alignment Scores:  
Pred. No.: 6,62e-257 Length: 1387  
Score: 2090.50 Matches: 412  
Percent Similarity: 93.85% Conservative: 0  
Best Local Similarity: 93.85% Mismatches: 2  
Query Match: 98.38% Indels: 25  
DB: Gaps: 1  
US-10-649-273-2 (1-414) x US-10-067-443-21 (1-1387)  
QY 1 MetLeuIleLeuThrIleuThraAlaGlyValPhePheIysProSerIysaArgLysValTyr 20  
DB 24 ATGCTTAATCTTGAAGTAACTGACGAGAGTTTCTTTTAAACCATCAAAAGAAAGTTTAT 83  
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrIleuPheLeuHisIleValIleu 40  
DB 84 GAATTTTAAAGAAAGTTTAAATTTTCAATCCGGAACACTATTTCTTCAAAAATGTATG 143  
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaIleValIleAspGluThrGlyAsnVal 60  
DB 144 GAAATTAATAATAATAAGAAAGAAAGAGGATTTGAGAGGGGCAATCTGTCT 203  
QY 61 LeuGlyValIleAlaIleHisSerGlnThrGluValHisLeuIysThrGlyIleValPro 80  
DB 204 TTGGAGAGAGCAATACATTCCTCCAAAGTGAAGTTTAAAAAACAGGGGAGATTTCT 263  
QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
DB 264 CCGAGAGCTCAACAGCTTCAAGAGAAATTCACAGAAATGATACAGAGAGCTTTCT 323  
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrTrlIleYsProGlyLeu 120  
DB 324 GCCAGTGAAGTCTCTCCAGTGAAGCTCTCAAGATTTGCAATTCATCAATCAAAACAGAGCTT 383  
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuIleLys 140  
DB 384 GCTTTAAGCTCGGAGAGGCTTATCATTTTATGATTCATGCTGTAGAGACAGTATTAATAAG 443  
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160

```

Db      444 CCATTCATCCCATCATCATATGAGAGCTCATGCACTTACTATTAGTTCAGCAAAATAA 503
Qy      161 ValGluPhenProphetLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuValLeuVal 180
Db      504 GTAGAAATTCCTTTTATGTTAGTCTTTTATGTTAGTCTTTTATGTTAGTCTTTTATGTTAGT 563
Qy      181 GInGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
Db      564 CAAGAGTTTCATATTTTCGCTTCTTGAAAGCTTTTGACATAGCAACAGGAGATG 623
Qy      201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyLysSerThrMetSer 220
Db      624 CTTGACAGGTGGCAAGAGACCTTTCTTTAATAAATCATCCAGAGTCTCCACCATGAGT 683
Qy      221 GlyGlyLysValAlaLeuGlyHisLeuValLysGlnGlyLysSerPheHisAspAspIleLys 240
Db      684 GGTGGAGAAAGCCATGAGCATTTGGCCAAACAGAAATGATTTTCATTTTGACATCAAA 743
Qy      241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
Db      744 CTTCCCTTGATCATGCTATAAAATTTGATTTTCTTTTACTGACATTCACACGTTACT 803
Qy      261 AspLysIleIleMetLysLysGlyLysGlnGlyLysIle----- 273
Db      804 GATTAATATATATGAAAAAGAAAAAGAGAGATATTTCTAATTAAGTAAAGTTGAA 863
Qy      274 -----GluLys 275
Db      864 CAGATAAATATTCCTGATGCTGCTTAAATAATACCTGCTCATTTCTGCAAGTATGAGAG 923
Qy      276 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 295
Db      924 GGGCAATCTGCTTCAGACAGACATGCTGCTCCACAGTACAGACACATGACATGCACT 983
Qy      296 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 315
Db      984 CATCTTGAAAAAGACACATCGGGCTATTTCTGTTTGTGTAAGCAGAGAGACTTGTTACT 1043
Qy      316 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArg 335
Db      1044 CAATAATATCAGTACGTGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1103
Qy      336 AlaLeuGlnIleLeuThrAsnAlaIleThrGlnCysThrLeuLeuCysProProProArgLeu 355
Db      1104 GCTCTGAAATTTTAAACAAAGCAACAGTCACTTTGTTGCTCTCCCAAGACTA 1163
Qy      356 CysThrAspAsnGlyIleMetIleAlaIleAsnGlyIleGluArgLeuArgAlaGlyLeu 375
Db      1164 TGCATCATATGATGATTAATGATTCATGATGATGATTAAGAACTACGCTGCTGCTGCT 1223
Qy      376 GlyIleLeuHisAspIleGlyGlyIleArgGlyGlyProLysCysProLeuGlyValAsp 395
Db      1224 GGCATTTTACATGATGATGAAGGATCGCTATGATGAACCAAAATGCTCTTGAGTAAAC 1283
Qy      396 IleSerLysGluValGlyGlyAlaSerIleLysValProGlnLeuLysMetGluIle 414
Db      1284 ATATCAAAAAGAGTTGAGAAAGCTTCATTAAGTACACACATTAATAATGAGATA 1340

```

```

; APPLICANT: Xue, Aidong J.
; APPLICANT: Weinman, Tom
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Dunrui
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 802
; CURRENT APPLICATION NUMBER: US/09/774,528
; CURRENT FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: PC_FL_genes Version 2.0
; SEQ ID NO 177
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (205)..(1305)
; US-09-774-528-177

Alignment Scores:
Pred. No.: 6,45e-213 Length: 1416
Score: 1747.00 Matches: 340
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.71% Mismatches: 0
Query Match: 82.21% Indels: 0
DB: 4 Gaps: 0

US-10-649-273-2 (1-414) x US-09-774-528-177 (1-1416)
Qy      74 LysThrGlyGlyIleValProProAlaLeuGlnGlnLeuHisArgGlyLysAsnIleGlnArg 93
Db      280 AGACAGGTGGATTTCTTCACAGCTCAACAGCTTCACAGAGAAATTTTCAACA 339
Qy      94 IleValGlnGlnAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAla 113
Db      340 ATAGTACAGAAAGCTTTCTGCGAGTGAAGTCTCCAAAGTACCTTCACAAATGCA 399
Qy      114 ThrThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGln 133
Db      400 ACTACCATTAACCAAGACTTCTTAAAGCTGGAAGTGGCTTATCATTTAGCTTACAG 459
Qy      134 LeuValGlyGlnLeuLysLeuProPheIleProIleHisMetGlnAlaHisAlaLeu 153
Db      460 CTGTAAGACAGTTAAAGCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 519
Qy      154 ThrIleArgLeuThrAsnLysValGluPheProPheLeuValLeuLeuIleSerGlyGly 173
Db      520 ACTATTAGGTGACCAATTAAGTAAATTTCTTTTATGTTCTTTTATGTTCTTGAGGT 579
Qy      174 HisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeu 193
Db      580 CACTGCTGTTGGCATTAAGTCAAGAGTTTCAGATTTTCTGCTCTTGGAAGTCTTTG 639
Qy      194 AspIleAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHis 213
Db      640 GACATAGACACAGGAGATGCTTGAAGAGGAGCAAGAGACTTTCTTTAATAAACAAT 699
Qy      214 ProGlyLysSerThrMetSerGlyGlyLysValAlaIleGlnHisLeuAlaLysGlnLys 233
Db      700 CCAAGAGTCTCCACATAGTGTGGAGAAAGCATTAAGCAATTTTGCCAAACAGAAAT 759
Qy      234 ArgPheHisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPhe 253
Db      760 AGATTTCAATTTGATGATCAACCTCCCTTGATCATGCTAATAAATTTGATTTCTTTT 819
Qy      254 ThrGlyLeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyLysIle 273
Db      820 ACTGACTTCAACACAGTACTGATTAATAATTAAGAAAAAGAAAAAGAGAGAGTATT 879
Qy      274 GlyLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMet 293

```

```

RESULT 3
US-09-774-528-177
; Sequence 177, Application US/09774528
; Patent No. 6743619
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhang, Jie
; APPLICANT: Zhao, Qing A.
; APPLICANT: Yang, Yonghong

```

```

Db      880 GAGAGGGGCAAACTCTGCTTCAAGCAGCAGACATTGCTGCCACAGTACAGACACAATG 939
Qy      294 AlaCysHisLeuValIysArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeu 313
Db      940 GCATGTCATCTTGTGAAAGAACACATCGGCTATTCCTTTTGTAGCAGAGACTTG 999
Qy      314 LeuProGlnAsnAsnAlaValIleuValAlaSerGlyIleValAlaSerAsnPheTyrIle 333
Db      1000 TTACCTCAAAATATATGACGATGCTGTCATCTGCTGTGTGTCGCAAGTAACTTCAATC 1059
Qy      334 ArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProPro 353
Db      1060 CGAGAGCTCTGAAATTTTAAACAAGCAACACAGTGCACCTTGTGTGCTCTCTCC 1119
Qy      354 ArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGlnArgLeuArgAla 373
Db      1120 AGCTATGCACTGTAATGCGATTATATGATGCAAGATGATGTAAGACCTACGTCCT 1179
Qy      374 GlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGluProIysCysProLeuGly 393
Db      1180 GGGCTGGGCACTTTTACATGACATGAGGAGCATCGCTATGAAACCAAAATGCTCTTGA 1239
Qy      394 ValAspIleSerIysGluValGlyGluAlaSerIleIysValProGlnLeuIysMetGlu 413
Db      1240 GTAGACATATCAAAAGAAAGTTGGAAGCTTCCATTAAGTACCAATTAAAAATGGAG 1299
Qy      414 Ile 414
Db      1300 ATA 1302

RESULT 4
US-10-067-443-23
; Sequence 23, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-23

Alignment Scores:
Pred. No.: 1,62e-166 Length: 1526
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 65.18% Indels: 0
DB: 4 Gaps: 0

US-10-649-273-2 (1-414) x US-10-067-443-23 (1-1526)
Qy      148 MetGlnAlaHisAlaLeuThrIleArgLeuThrAsnIysValGlnPheProPheLeuVal 167
Db      1 ATGAGGCGCTCATCACTTACTATAGGTGACCAATAAAGTAGAATTTCTTTTACTT 60
Qy      168 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 187
Db      61 CTTTGAATTTCTGAGAGCTCACTGCTGTGGCATTAGTTCAAGAGTTTCAGATTTCCTG 120
Qy      188 LeuLeuGlyIysSerLeuAspIleAlaProGlyAspMetLeuAspIysValAlaArgArg 207

```

```

Db      121 CTTTGGAAAGCTTTTGACATAGCACACAGGTGACATGCTTGACAGGTGCGAAGAGA 180
Qy      208 LeuSerLeuIleIysHisProGluCysSerThrMetSerGlyIleYsaAlaIleGluHis 227
Db      181 CTTTCTTTAATTAACATCCAGAGTGTCCACCAATGAGTGGTGGGAAAGCCATAGACAT 240
Qy      228 LeuAlaIysGlnIysAsnArgPheHisPheAspIleIysProProLeuHisHisAlaIys 247
Db      241 TTGGCAAAACAAGAAATGATATTTTATTTGACATCAAACTCCCTTGATCATGTGTAA 300
Qy      248 AsnCysAspPheSerPheThrGlyLeuGlnHisValIleThrAspIleIleIleLeuIys 267
Db      301 AATTGTGATTTTCTTTTATCTGACCTTCAACACGTTACGATTAATAATATGAAAAAG 360
Qy      268 GluIysGlnGluGlyIleGluIysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIle 287
Db      361 GAAAAGAGGANGATTTGAGAGGGGCAAAATCTGTCTTACGACGACACATTCGTGCT 420
Qy      288 ThrValGlnHisThrMetAlaCysHisLeuValIysArgThrHisArgAlaIleLeuPhe 307
Db      421 ACAGTACAGCACACATGCGATGTCATCTGTGAAAAGAACACATCGGGCTATTCGTGTT 480
Qy      308 CysIysGlnArgAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGlyIleVal 327
Db      481 TGTAAAGCAGAGACTGTGTACTTCAAAATATGAGTACGTGTTGCATCTGTGTGTGTC 540
Qy      328 AlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 347
Db      541 GCAAGTAACTTATATATCCGACAGACTCTGGAATTTTAAACAAGCAACACAGTGCAT 600
Qy      348 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 367
Db      601 TTGTTGTGTCCTCCCTCCAGACTATGCACTGATTAATGCGATTATGTCAGATGAT 660
Qy      368 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 387
Db      661 ATTGAAAGCATACGTCGTGGCTTGAGCATTTTACATGACATGAGAGCATCGCTATGAA 720
Qy      388 ProIysCysProLeuGlyValAspIleSerIysGluValGlyAlaSerIleIysVal 407
Db      721 CCAAAATGTCCTCTTGAGTACATATCAAAAGAGTTGGAGAGCTTCCATTAAGTA 780
Qy      408 ProGlnLeuIysMetGluIle 414
Db      781 CCACAAITTAATAATGACATA 801

RESULT 5
US-10-067-443-20
; Sequence 20, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 14364
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-20

Alignment Scores:
Pred. No.: 2.19e-139 Length: 14364
Score: 1186.50 Matches: 313
Percent Similarity: 32.30% Conservative: 1

```

Best Local Similarity: 32.20% Mismatches: 2  
Query Match: 55.84% Indels: 657  
DB: 4 Gaps: 4  
US-10-649-273-2 (1-414) x US-10-067-443-20 (1-14364)

QY 74 LysThrGlyGlyIleValProProAlaAlaGlnIleuHISarGluAsnIleGlnArg 93  
DB 10623 AGAAGCGTGGATGTTGCTCCAGCAGCTCAACAGCTTCACAGAGAAATATTCAGCA 10682

QY 94 IleValGlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAla 113  
DB 10683 ATAGTCAGAGAACCTCTTCTGCGAGTGTCTCCAGTGCCTCTCCAGCAATTCGA 10742

QY 114 ThrThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGln 133  
DB 10743 ACTACCATTAACACGAGCTTCTTMACTGGAGTGGAGCTTATCATTTAGCTTACAG 10802

QY 134 LeuValGlyGlnLeuLysLysProPheIleProIleHISHISMeGluAlaHISAlaLeu 153  
DB 10803 CTGGTAGACAGTTAAABAAAGCCATTCATCCCATCATATGAGGCTCATGCACTT 10862

QY 154 ThrIleArgLeuThrAsnLysValGluPheProPheLeuValIleuLeuIleSerGlyGly 173  
DB 10863 ACTATTAGGTGGACCATTAAGTAGATTCTTTTATGTTCTTTGATTCTGGAGGT 10922

QY 174 HISCyLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeu 193  
DB 10923 CACTGCTGTGGCATTAAGTTCAGAGAGTTCAGATTTCGCTTCATGGAAAGTCTTG 10982

QY 194 AspIleAlaProGlyAspMetLeuAspLysVal----- 204  
DB 10983 GACATGACACGAGTGACATGCTTGACAGAGT-AAATTAAGATTAAATTTCTCCATTCTT 11041

QY 204 ----- 204  
DB 11042 TTGTATGTTGCCATTTCAACTAAGTAGCAATGATGTCTACCACTTACCTAAA 11101

QY 204 ----- 204  
DB 11102 TATTTCTGAATTTATCTTAGTAACAGAAAAATTACATATGTGTGAGAAAAATAGA 11161

QY 204 ----- 204  
DB 11162 AAGATGATACAAATTTATATTTCTTACCTTTCTTAATAAATGTAAAGGTTTCT 11221

QY 204 ----- 204  
DB 11222 ATCTGATCATAAAGCTGAATTAAGTTCAGATACAGTATGATTTTGCATAATATGT 11281

QY 204 ----- 204  
DB 11282 ATGTGAAGAAGCTTCTGTAACATACTGCAAAAAAGGTAAATTAAGAAATAT 11341

QY 204 ----- 204  
DB 11342 ATATAGATTAAACATTAAGACATTAAGATGCAAGATTAATCAACAATTAATCT 11401

QY 204 ----- 204  
DB 11402 TACACCAACAGAGGTCCTCCCTCTTTGTTTGAATACTACAGAGGTAATCT 11461

QY 204 ----- 204  
DB 11462 GCATATATAGAAAAACAACAACAACAACAACAACAACAACAACAACAACAACA 11521

QY 204 ----- 204  
DB 11522 TAGAGATATAGCAAGATCTTATTAATGAAGTTCAATTAAGAGATTAATGTCACT 11581

QY 204 ----- 204  
DB 11582 TCAGCCCATTTTCCAAACCAATAGAGCAACAATAGACAGGGGAGTGAATGGCTCT 11641

QY 204 ----- 204  
DB 11642 TATTTGCGGTCAATTAAGAACAGGGTGTCTGCTTACCTGAATACAGTAATGTC 11701

QY 204 ----- 204  
DB 11702 TATATTTGCCAAGATTAAGCATGTTTATTCATTACAGGGGTTTTTTGTTTGTAGTAAT 11761

QY 204 ----- 204  
DB 11762 TTCAATTTATTTCTTTCGATCTTTTCGTTTCACAGATTTAAATTTTATGACTAAA 11821

QY 205 ----- 205  
DB 11822 AATATGTTTCTTGATAGAGTGGCAAGAGACTTTCTTAAATTAACATCCAGAGTCTCC 11881

QY 218 ThrMetSerGlyGlyLysAlaIleGluHISLeuAlaLysGlnGlyAsnArgPheHISPhe 237  
DB 11882 ACCATGAGTGTGGAAAGCCATAGAACATTTGGCCAAACAGGAATAATGATTTTCATTTT 11941

QY 238 AspIleLysProProLeuHISHISAlaLysAsnCyAspPheSerPheThrGlyLeuGln 257  
DB 11942 GACATCAAACTCCCTTGATCATATGCTTAATAATGTGATTTTCTTTACTGACTTCA 12001

QY 258 HISValThrAspLysIleIleMeLysLysGluLysGlu----- 271  
DB 12002 CACGTTACTGATTAATAATATGAAAAAGGAAGGATATTTCTAATTAAGT 12061

QY 272 ----- 272  
DB 12062 AAGTTGAACAGATTAATATCTGATGTTGCCATAAATAGCTCATTTCTCGACG 12121

QY 272 YIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHISTh 292  
DB 12122 TATTGAGAGGGGCAAACTCTGCTTCAGACAGACATTCCTGCAAGATACAGACAC 12181

QY 292 IMetAlaCyHISLeuValLysArgSThrHISArgAlaIleLeuPheCyLysGlnArgAs 312  
DB 12182 AATGCAATGATCTTGTGAAAAGAACACATCGGCTATTCTGTTTGTAAAGAGAGCA 12241

QY 312 PLeuLeuProGlnAsnAsnAlaValLeu----- 321  
DB 12242 CTTGTTACTCAAAATATATGACGATCTGTAAATTTATCTCATTTATAGTAATAGTTA 12301

QY 321 ----- 321  
DB 12302 CACTTGCATATATGTTACTTTTTCACAAAGCTTGACCTTGTTGTAGATGAACAGAT 12361

QY 321 ----- 321  
DB 12362 CTTTATGCTTATAGTACCCCTGACAGATGAATAATTAAGCATAGAAAGCTTAACAG 12421

QY 321 ----- 321  
DB 12422 CCATTTCTTGACTAGTTTGTAGCTTATAGGACAGCTGTATAGCTTCTATAGCACATA 12481

QY 321 ----- 321  
DB 12482 AGTCTAATTTTGATCTTCTGTGTGATTTAAAGAGGCTTAACAATTAAGAAAGTAAT 12541

QY 321 ----- 321  
DB 12542 GCAGTAAGTGTATCACTATTTTGAAGAAAAATAGGTGATTTCTTCATCTTTGATGA 12601

QY 321 ----- 321  
DB 12602 ATCCCTTGTGTTGTTGTTTATTAATAGCAGTCAATTTAGCAGTGGAGGTGAT 12661

QY 321 ----- 321  
DB 12662 TCCAACTTGTGACACTAATGTTGATTAAGTTCTGATTAATCACTATATTTGACAGCC 12721

```
QY 321 ----- 321
Db 12722 AAAATCCCTTAATGCTTAAAGCCTTGACAAACATCCCTTTAACTGATCTTAAA 12781
QY 321 ----- 321
Db 12782 CTTTATTCATTTAAAAATTATAAATAAGTGGAAAAATTAAATGATGATTAATTCAT 12841
QY 321 ----- 321
Db 12842 AGATGGAATTTTACATGATATCAAGAAATATTTTTCAGAGTATGTAATAAATGCA 12901
QY 321 ----- 321
Db 12902 CAAAATAATAAAATTTCAAGGCTTAATAATAGTGTACTGATGTAATTAATAATA 12961
QY 321 ----- 321
Db 12962 AATATTAGATGAAGGTTGGAAGAAATATACAAAAATGCTAGTAATGTTGTATGCTA 13021
QY 321 ----- 321
Db 13022 TTAGATTAATTAATTTTCTTCCAAATTTTATTAATGATGATGATGATCTGCC 13081
QY 321 ----- 321
Db 13082 CATTAACCATCTCAAAATGGAGATGTTATTAATGTTAAATGATGATTTTCTCAGG 13141
QY 321 ----- 321
Db 13142 TTTAATTAGCAGCTGGTTCATATCCATATGATGATGATGTTTGTCTCAATTCCT 13201
QY 322 -----ValAlaSerGlyValAlaSerAspPheTyrIleArgArgAlaLeuGluIle 340
Db 13202 TCAGGTTGCACTGGTGGTGTGCGAAGTAACTTATATGCGAGACCTCGGAAATTTT 13261
QY 340 uThrAsnaIaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnG 360
Db 13262 AACAAACGACACAGTGCATTTGTTGTCTCCCTCCAGACTAGCATGCTGATATGG 13321
QY 360 yIleMetIleAlaTyr----- 365
Db 13322 CATTAATGATGACGTGTAAGCCACAGAGATATAGCTGCTCATCATTAATGTAATAT 13381
QY 365 ----- 365
Db 13382 TAATTGCCATTTTATCATTAAGCTTCTTCTTCAGATCTTGAGCTTATGATTTTAT 13441
QY 366 -----AsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuH 379
Db 13442 TTTAATGCTTCTTATTTAGAAATGTAATTAAGACTACGCTGCTGGCATTTTAC 13501
QY 379 IaAspIleGluGlyIleArgTyrGluProLys 389
Db 13502 ATGACATAGAAAGCATCCGCTATGAACCAAG 13553

RESULT 6
US-09-596-002-38
; Sequence 38, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Patterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; PRIOR FILING DATE: 2000-05-16
; PRIOR APPLICATION NUMBER: 60/140,121
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 38
```

```
; LENGTH: 94750
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURES:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyre template ID No. 6632636 38
; PUBLICATION INFORMATION:
US-09-596-002-38

Alignment Scores:
Pred. No.: 2,66-50 Length: 94750
Score: 502.00 Matches: 129
Percent Similarity: 50.81% Conservative: 59
Best Local Similarity: 34.86% Mismatches: 149
Query Match: 23.62% Indels: 34
DB: 4 Gaps: 9

US-10-649-273-2 (1-414) x US-09-596-002-38 (1-94750)
QY 39 ValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValAlaSerGluThr--- 57
Db 23873 GTATTGGATTGGAGACATCTTGATGAAACAGGGCTTGCATCTATGATGATGATG 23932
QY 58 -----GlyAsnValLeuGlyGluAlaIleHisSerGlnThrGluValHisLeu 73
Db 23933 AATGGCGTGGTGGCGGCGCTGTCTCAGGTTTGTATACCAAAATTAATCTATCATGCC 23992
QY 74 LysThrGlyIleValProProAlaGlnGlnLeuHisArgValAsnIleGlnArg 93
Db 23993 ACTTATGCGGTGTCGTGCTGACCTTGACAGTGCAGACCAATTCGTAACCTTGTGCCG 24052
QY 94 IleValGlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAla 113
Db 24053 TTATTTAATGATGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATG 24112
QY 114 ThrThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGln 133
Db 24113 TATACCAAGGCGCAGGCGTATGGGGCGTGTATGACAGGGCATTTTGGGGCGAG 24172
QY 134 LeuValGlyLeuLysPheProPheIleProIleHisIleMetGluAlaHisAlaLeu 153
Db 24173 CTGGCGTATGGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGG 24232
QY 154 ThrIleArgLeuThr-----AsnLysValGluPheProPheLeuValLeuLeuIleSer 171
Db 24233 GCACCGCTATTTGGCAGTATGACCCCAAGCTTTCCGTTGTGTGCTGCTGCTGCTG 24292
QY 172 GlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLys 191
Db 24293 GGGCGTATACCAATGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 24352
QY 192 SerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIle 211
Db 24353 TCTATCGATGATGCGGTGGTGAATGCTTGTATAAACGCAAAATATGCTCAAACTG--- 24409
QY 212 LysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGln 231
Db 24410 CCTATCT-----GTTGGCCCAATATGCAAAATTAATGCAAAAC 24451
QY 232 GlyAsnArgPheHisPheAspIleLysProProLeuHisIleAlaLysAspCysAspPhe 251
Db 24452 GGCACCCACACCGCTATGACCTGCCAAGCCATGACGAT---AAAGGCTGATTTT 24508
QY 252 SerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLysGlyLysGlu 271
Db 24509 TCGTTTCAGTGCAGAAACCGCATTCATATCTCAATAAGACACACCAAGCCCA 24568
QY 272 GlyIleLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHis 291
Db 24569 AGCGACCC-----GCCACAGACAGACATCGCCGCAAGCTTTGAGTAT 24613
QY 292 ThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArg 311
```

```

Db      24614 GCGGTGGTGAATCTTGTGCAAAAATGACCAACACTACATGACAGCATTCC 24673
Qy      312 AspleuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsn 331
Db      24674 CAG-----CTGGTGTGCGACGGGGCGCTCTGCCAATCAG 24709
Qy      332 TyrLeuArgAlaLeuGluLeuThrAsnAlaThrGlnCysThrLeuLeuCysPro 351
Db      24710 ATGCTACGCCGACCTGACCCGACGCTCCGCAATCGATCGCTGGTACTATGCC 24769
Qy      352 ProProArgLeuLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeu 371
Db      24770 CCGACCGACCTATGACCGATATATGCGCATGATCGCTGATGCGCTTTTGTGGCTC 24829
Qy      372 ArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgGlyProGlyCysPro 391
Db      24830 AGCTGTGGACATCG-----GATGACTTGGC-GGTTGCTGTATTC-----CCG 24873
Qy      392 LeuGlyValAspIleSerGlyValGly 401
Db      24874 ATGGGATATGACGACGCTTGGCCTATCGGC 24903

RESULT 7
US-09-540-236-806
; Sequence 806, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 806
; LENGTH: 1053
; TYPE: DNA
; ORGANISM: M. catarrhalis
US-09-540-236-806

```

```

Alignment Scores:
Pred. No.: 1,236-53 Length: 1053
Score: 501.00 Matches: 126
Percent Similarity: 49.86% Conservative: 56
Best Local Similarity: 34.52% Mismatches: 139
Query Match: 23.58% Indels: 44
DB: Caps: 8

```

US-10-649-273-2 (1-414) x US-09-540-236-806 (1-1053)

```

Qy      39 ValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThr-- 57
Db      16 GTATGGGATGAGACATCTTGATGAAACAGGGCTGCATCTATGATAGATGATG 75
Qy      58 -----GlyAsnValLeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLys 74
Db      76 AATGGACGGGGCGCGCTGCTGCTCAGGTTTATATGATCAATTAATTAACATGACCA 135
Qy      75 ThrGlyIleValAlaProProAlaAlaGlnGluLeuHisArgGluAsnIleGlnArgIle 94
Db      136 TATGGGGTGTCTGCTGAGCTTGACCTGCGACGACCACTTGGTAAGCTTGCCGTTA 195
Qy      95 ValGlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThr 114
Db      196 TTTAATGAGATGTTGATGACGCAAAATATCACAAATCCGACATGATGCGGTGGCGTAT 255
Qy      115 ThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeu 134
Db      256 ACCAAAGACCCCGCGGCTGATGGGCAATGATGACAGGGCGCATATTTGGCGGACGCTG 315
Qy      135 ValGlyGlnLeuLysLysProPheIleProIleHisMetGluAlaHisAlaLeuThr 154
Db      316 GCCTATGAGCTGGCGCTGACGCGTTGGGGTGATCATATATGAGGACATCTGTAGCA 375

```

```

Qy      155 IleArgLeuThr-----AsnLysValGluPheProPheLeuValLeuLeuIleSerGly 172
Db      376 CCGCTATTGGCCAGTGTATGACCCCCCAAGCTTTCGTTTGTGTGCTGCTGTCGGGCG 435
Qy      173 GlyHisCysLeuLeuAlaLeuValGlnGlyValAsnAspPheLeuLeuGlyLysSer 192
Db      436 GGTATACCATCATGCTGGTGGCTGCGATGTGGGCGGTGATACAGATATTTGGCGAGTCC 495
Qy      193 LeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuLys 212
Db      496 ATCATATGATGCGGTGGCGGAATGCTTGATTAACGCAAAATCTCAAACTG---CCC 552
Qy      213 HisProGluCysSerThrMetSerGlyValLysAlaIleGluHisLeuAlaLysGlnGly 232
Db      553 TATCTCT-----GGTGTCCAATATCGAAATAATGACCAAAAACGCG 594
Qy      233 AsnArgPheHisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSer 252
Db      595 AACCCACACGCGCTATGACCTGCCAAGACCGATGACGAT---AAGGCGCTGATTTTTCG 651
Qy      253 PheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGly 272
Db      652 TTGATGGCATGAATAACCGCATTCATATATCTCAATAAGACACGCCAAAGCCAAAGC 711
Qy      273 IleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThr 292
Db      712 GACCT-----GCCACACGACGACATCGCGCAAGCTTGATGATATCG 756
Qy      293 MetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLeuGlnArgAsp 312
Db      757 GTGTGATGATCTTGGTCAAAAATGACACACGACATACAGATGACAGCATTCGCGCAG 816
Qy      313 LeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsnPheThr 332
Db      817 -----CTGTGTGTGCGAGGGGCGTCTGTCACATACAGC 852
Qy      333 IleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProPro 352
Db      853 CTACGCCGACCTGACCGGACGACGCTGCGCAAAATCGATGCGTGTACTACGCCCGCG 912
Qy      353 ProArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArg 372
Db      913 ACCGAGCATATCAGGATATGATGCGATGATCGCTTACGCTTTTGTGGCTAAGC 972
Qy      373 AlaGly-----Leu 375
Db      973 CGTGGCAGTCGATGACTTGGCGGTGCTGTATTTCCCGATGGATATGACATGCTT 1032
Qy      376 GlyIleLeuHisAsp 380
Db      1033 GGTATCGAATATGAT 1047

RESULT 8
US-09-252-991A-801
; Sequence 801, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,768
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 801
; LENGTH: 1206
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa

```

US-09-252-991A-801

## Alignment Scores:

Pred. No.: 2,43e-53 Length: 1206  
 Score: 499.50 Matches: 133  
 Percent Similarity: 51.68% Conservative: 67  
 Best Local Similarity: 34.37% Mismatches: 158  
 Query Match: 23.51% Indels: 29  
 DB: 4 Gaps: 10

US-10-649-273-2 (1-414) x US-09-252-991A-801 (1-1206)

```

QY 13 LysProSerLeuArgLysValTyrGluPheLeuArgSerPheAsnPheHisProGlyThr 32
DB 91 AAAGCCTCCGATCGAATAACACTGCTGTGCGCAAGCGCCGCAATATATACGCGGCTT 150
QY 33 LeuPheLeuHisLysLe-----ValLeuGlyLeuGlyThrSerCys 46
DB 151 CGTATTTGCCAAGCTTAAAGCTCAAGCCCAATGCGGTGCTGGAAGTGCCTCTGC 210
QY 47 AsnAspThrAlaAlaAlaValAlaValAspGluThrGlyAsnValLeuGlyAlaAlaHis 66
DB 211 GACGAAACCGGGGCTGCCCTTTACGACAGCAAGCGGCTGCTGCGCCAGCGCTTTC 270
QY 67 SerGlnThrGluValHisLeuLysThrGlyGlyLeuValProProAlaAlaGlnGlnLeu 86
DB 271 AGTCAGATGACCTCATCGCTCATCGGCGGCGGTGCTGCGCGAGCTGCGCGGAC 330
QY 87 HisArgGluAsnLleGlnArgLleValGlnGluAlaLeuSerLysAspGlyValSerPro 106
DB 331 CACGTCAACGGATGCTGCCGCTGATCCGCAAGCTCTCAACAGTCCGCTGCACGCCG 390
QY 107 SerAspLeuSerAlaAlaAlaThrThrLysProGlyLeuAlaLeuSerLeuGlyVal 126
DB 391 GGGATATGACGCGCATCCCTATACCGCGGCTCCGCTGCTGCGCGCTGCTGTG 450
QY 127 GlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLysProPheLleProLleHis 146
DB 451 GGGGCTTCTGTGTCGACGAGTGGCTTGCCTGCGGCGGTGCGCGGCTGCGCAC 510
QY 147 HisMetGluAlaHisAlaLeuThrLleArgLeuThrAsnLys--ValGluPheProPhe 165
DB 511 CACATGGAAGGCGCATCTGCGCGCATGCTGGAAGAGACGACCGCGGTTC 570
QY 166 LeuValLeuLeuLleSerGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAsp 185
DB 571 GTCGCTTCTGTGTTTCCGCGGTGACACCCAGTGTGCGGCTGACGATATCGGCCG 630
QY 186 PheLeuLeuLeuGlyLysSerLeuAspLleAlaProGlyAsnMetLeuAspLysValAla 205
DB 631 TACCAAGTTGCTTGGCGCATCGGTGACGATGCCGCGGGAACCTTCGACAAAGCCGC 690
QY 206 ArgArgLeuSerLeuLleLysHisProGluCysSerThrMetSerGlyLysAlaLle 225
DB 691 AAGCTGATCGGCTG---GGCTATCCC-----GGTGTCCGGAATC 729
QY 226 GluHisLleAlaLysGlnGlyAsnArgPheHisPheAspLleLysProProLleHisHis 245
DB 730 GCCCGCTGCGGAGCGGCACTCTGCGCGCTTGTCCCGCGCGGATACCGCAT 789
QY 246 AlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysLleLleMet 265
DB 790 CGCCCGCGCTGACCTTCACGCTTCACGCGGCTCAACACCTTACCTCAAGC---ACCTGG 846
QY 266 LysLysGluLysGluGlnGlyLleGlyLysGlyGlnLleLeuSerSerAlaAlaAspLle 285
DB 847 CAGCGTGTGCTGAGGCGCGGCGACGACGAGCAG-----ACCCGCTCGACATC 897
QY 286 AlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaLle 305
DB 898 GGCCTGCGCTTCAGACCGCGGTGTCGACACCTCTGATCAAGTGGCGCTCGCGCTTC 957
QY 306 LeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGly 325
  
```

```

DB 958 -----AAGCAGCCGCGCTG-----AAGAAC-----CTGTATATGCCGCGC 993
QY 326 GlyValAlaSerAsnPheThrLleArgArgAlaLeuGlyLleLeuThrAsnAlaThrGln 345
DB 994 GTGTGACGCCCAACAGAGGCGCTGCGAGGCGCTGGAATAAGATGCGGAATAGAG 1053
QY 346 CysThrLeuLeuCysProProProArgLeuCysThrAspAsnGlyLleMetLleAlaTrp 365
DB 1054 GGGCAGGTGTTCATACGCGCGCGCTGCTGCACCGACATGCGCGGATGATGCGCTTAC 1113
QY 366 AsnGlyLleGluArgLysAlaGlyLeuGlyLleLeuHisAspLleGlnGlyLleArg 385
DB 1114 GCGCGCTGCGACGCGCTGCTGCGCGC-----CAGCATGACGCGCGCGGATACG 1164
QY 386 TyrGluProLysCysProLeu 392
DB 1165 GTCCAGCGCGCTGCGCGGATG 1185
  
```

## RESULT 9

```

US-09-252-991A-884/c
; Sequence 884, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 884
; LENGTH: 1059
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-884
  
```

Alignment Scores: 1.52e-52 Length: 1059  
 Pred. No.: 492.50 Matches: 128  
 Score: 53.24% Conservative: 61  
 Best Local Similarity: 36.06% Mismatches: 143  
 Query Match: 23.18% Indels: 23  
 DB: 4 Gaps: 9

US-10-649-273-2 (1-414) x US-09-252-991A-884 (1-1059)

```

QY 39 ValLeuGlyLleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThrGly 58
DB 1038 GTGTGGAAGTGTGAAGCTCTTGACGAAACCGGCTGCTTTCGACGAGCAAGCGC 979
QY 59 AsnValLeuGlyGluAlaLleHisSerGlnThrGluValHisLeuLysThrGlyGlyLle 78
DB 978 GGCCTGCTGCGCGACGCGGCTTTCATGACATGCACTCCATCCGCTACGCGGCGCTC 919
QY 79 ValProProAlaAlaGlnGlnLeuHisArgGluAsnLleGlnArgLleValGlnGlnAla 98
DB 918 GTGCGCGAGCTGCGCTGCGCGGACACGTCAGAGGAGTGTGCGCGATGCGCGAGGTG 859
QY 99 LeuSerAlaSerGlyValSerProSerAspLeuSerAlaLleAlaThrThrLleLysPro 118
DB 858 CTCGACGAGTCCGCTCATCGCGCGCGGATATCAACCGCATGCGCTATACCGCGCTTC 799
QY 119 GlyLeuAlaLeuSerLeuGlyValGlyLysSerPheSerLeuGlnLeuValGlyGlnLeu 138
DB 798 GGCCTGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 739
QY 139 LysLysProPheLleProLleHisHisMetGluAlaHisAlaLeuThrLleArgLeuThr 158
  
```





Db 946 ACCGATATATGTCCTTGGCGGAGTATCCGTTTAAAGTGTACCGAG 1005  
Qy 376 GYILEUHIHISAPILLEGUIGLYIARGTYRGLUPROBYSYSPROLEU 392  
Db 1006 GGGCCTTTA-----GGGTGACAGTGAAGACAGCTTGCCCTTTA 1044  
RESULT 11  
US-09-902-540-6612  
Sequence 6612, Application US/09902540  
Patent No. 6833447  
GENERAL INFORMATION:  
APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Wiegand, Roger C.  
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-10(15849)B  
CURRENT APPLICATION NUMBER: US/09/902,540  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/217,883  
PRIOR FILING DATE: 2000-07-10  
NUMBER OF SEQ ID NOS: 16825  
SEQ ID NO 6612  
LENGTH: 996  
TYPE: DNA  
ORGANISM: Myxococcus xanthus  
US-09-902-540-6612  
Alignment Scores:  
Pred. No.: 3,01e-51 Length: 996  
Score: 482.00 Matches: 128  
Percent Similarity: 52.48% Conservative: 52  
Best Local Similarity: 37.32% Mismatches: 131  
Query Match: 22.68% Indels: 32  
DB: 4 Gaps: 9  
US-10-649-273-2 (1-414) x US-09-902-540-6612 (1-996)  
Qy 38 ILEVALLEUGLYILEGLUTHSERCYSPASPPTHRAIAIAIAVAIAASPGLUTHR 57  
Db 4 CTCGCTCTAGAGTGTGAACCTGTGTATGAGACTCCCGCGCTGTGAGAGAGCGC 63  
Qy 58 GIYASNAVLEUGLYIGLYIAIAIEHISSEGLINTHRLVIAHISLEUYSYTHGLY 77  
Db 64 CGCGCGCGCTGTGCGATGTCCTCAGCAGAGTGACATCCACGCGGTGGCGGG 123  
Qy 78 ILEVALPROBIAIAIAGLINTLEUHIHISARGLUNSLIEGLINAGLILEVALGLIN 97  
Db 124 GTGGTCCGAGCTGGCCAGCCCAACACATGCTCCAGGTGCGCGCTCCACGAG 183  
Qy 98 AIALEUSERALASERGLYALSERPROSERASPLEUSERALIAIAIAIAIAIAIAIA 117  
Db 184 GGGCTACCGCGGCGCAACAGACGCTGCAGACGTGACCTCAGCGCTCAGCGGC 243  
Qy 118 PROGLYLEUAIALAEUSERLEUGLYIGLYLEUSERPHEUSERLEUGLINTLEUVALGLIN 137  
Db 244 CCGGACTATCGCGCGCGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 303  
Qy 138 LEULYELYPROBHEILEPROLEHISHSIMETGLUAIHISALAEUTHRLIEARGLEU 157  
Db 304 ACGGCGAAGCCCTTCTGGGCGCAACACCTGAGAGGCGACCTGCGCATCCGCTG 363  
Qy 158 THRASLUYVAL---GLUPHEPROBHEUVALLEUULEULESERGLYGLYHISCYLEU 176  
Db 364 TTGAGAGTGGCGCGGAGCGCGCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 423  
Qy 177 LEUAIALAEUVALGNGLYVALSERASPBHEULEUSERLEUGLYYSERLEUASPIIA 196  
Db 424 CTCCTACGAGGTGAGCGCTTACGGGCGAGTACCGCTGTGTGGCGACGCGGAGCGG 483  
Qy 197 PROGLYASPMETLEUASPLYSVALAIARGARGLEUSERLEUULEYSHISPROGLUCYS 216  
Db 484 GCGGCGGAGGCAATATGACAAAGACCGCTCGATCTCGGCGCTG---CCGTATCCG----- 534

Qy 217 SERTHMETSERGLYLYSALAIIEGLUHISEUAIALYSGLINGLYASN----- 233  
Db 535 -----GGTGGCGAGCCCATGACAGTGTGGCGAGCGAGGCAACCCGAGGCC 582  
Qy 234 ---ARGPHEHISPHAEAPILIEYSPROBHEUHIHISIALYASPCYASPBHEISER 252  
Db 583 ATCCGCTTC-----CCGCGCGCGTGGCCGCGGCAACACTTGCAGCTGTCC 627  
Qy 253 PHETHRGLYLEUGLINTHISVALTHRASPLYSLEILEMETLYSLYGLUYGLUGLUGLY 272  
Db 628 TTCTCCGGGTGAAG-----ACGCGGTGCTGCACCACTGACAGACAGCGC 675  
Qy 273 ILEGLUYSGLYINILEUSERSERALIAIASPILAEALIAIAIAIAIAIAIAIAIA 292  
Db 676 GTGCCGAGGCGGAGGCGCTG-----GCGATTGTGCGGCTTCACAGAGGCC 726  
Qy 293 METIACYSHISLEUVALYVARGTHRIARVALIAIELEUPHECYSLYGLINARGSP 312  
Db 727 GTGGCGACGTGCTGTCAAGAG-----CTGTGGCCCGCGCGCGCGG 771  
Qy 313 LEUEUPROGINAENASNAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIA 332  
Db 772 TTG-----GCCCAAGACACTTGTCTGTGCGCGCGCTGCGCGCGCGCGG 822  
Qy 333 ILEARGALAEUGLILEUITHRASNAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIA 352  
Db 823 CTGGCGGACACTGTGTCAAGCGCGAGCGAGGCGGGGCTTGAACATGCTTCGCGCG 882  
Qy 353 PROARGLEUCYTHRASPBNGLYILEMETIIEALIAIAIAIAIAIAIAIAIAIAIA 372  
Db 883 GTGGCGGTGTGACAGCAATATGCGCATATTCGCTGGCGGGGTATGAGCGTACCGC 942  
Qy 373 AIAGLYLEU 375  
Db 943 CGCGGCGCTG 951  
RESULT 12  
US-09-902-540-503/C  
Sequence 503, Application US/09902540  
Patent No. 6833447  
GENERAL INFORMATION:  
APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Wiegand, Roger C.  
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-10(15849)B  
CURRENT APPLICATION NUMBER: US/09/902,540  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/217,883  
PRIOR FILING DATE: 2000-07-10  
NUMBER OF SEQ ID NOS: 16825  
SEQ ID NO 503  
LENGTH: 2582  
TYPE: DNA  
ORGANISM: Myxococcus xanthus  
US-09-902-540-503  
Alignment Scores:  
Pred. No.: 1.62e-50 Length: 2582  
Score: 482.00 Matches: 128  
Percent Similarity: 52.48% Conservative: 52  
Best Local Similarity: 37.32% Mismatches: 131  
Query Match: 22.68% Indels: 32  
DB: 4 Gaps: 9  
US-10-649-273-2 (1-414) x US-09-902-540-503 (1-2582)  
Qy 38 ILEVALLEUGLYILEGLUTHSERCYSPASPPTHRAIAIAIAIAIAIAIAIAIAIA 57  
Db 1525 CTCGCTCTAGAGTGTGAACCTGTGTATGAGACTCCCGCGCTGTGAGAGAGCGC 1466

```

QY      58 GlyAsnValLeuGlyLalaIleHisSerGlnThrGluValHisLeuLysThrGlyGly 77
DB      1465 CCGCCGCGCGTTGGATGTCGTCCTCCACGAGTGACATCCAGCCGGGCGTGGGG 1406
QY      78 IleValProProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGlu 97
DB      1405 GTGGTGCACGAGCTGCACACCCGCAACCATCATCTGCACCGCCGCGTCCACGAG 1346
QY      98 AlaLeuSerAlaSerGlyValSerProSerAlaIleAlaThrThrIleLys 117
DB      1345 GGGCTACCGGGGCAACAGACGCTTCAGACGTGACCTCATCGCCGTCACGTCGGCC 1286
QY      118 ProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnValGlyGln 137
DB      1285 CCGGAGCTACGCGCGCGCTGCGTGGAGTGAGTGAGTGCACAGGCGCTGAGCGTGGCG 1226
QY      138 LeuLysLeuProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeu 157
DB      1225 ACCGCGAAGCCCTTCGCGGCGCCACCATCTGAGGCGCACCTGCGCATCCGCGCTG 1166
QY      158 ThrAsnLysVal---GluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeu 176
DB      1165 TTGGAGGTGGCGCGCGCGCGCTGCTGCGCTGCTGCTGCTGCTGCGCGGACACAGC 1106
QY      177 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 196
DB      1105 CTCTACGAGTGCAGGCGCTACGCGGACGTACCGGCTGTGGGACGACCGCGAGCGG 1046
QY      197 ProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCys 216
DB      1045 GCGCGGAGGACATATGACAGACCGCTCGCATCTCGCGCTG---CCGTATCCG----- 995
QY      217 SerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsn----- 233
DB      994 -----GGTGGCAGCCCATGACACAGTTCGCGGACGACGAGGAAACCGGAGGCC 947
QY      234 ---ArgPheHisPheAspIleLysProProLeuHisHisAlaLysAsnAspPheSer 252
DB      946 ATCCGCTTC-----CCGCGCGCGCTGCGCGGCGCAACATTGACGTGCTCC 902
QY      253 PheThrGlyLeuGlnHisValThrAspLysIleIleMetLysGlyLysGluGluGly 272
DB      901 TTCTCCGGGTGTAAG-----ACGCGCGTGTGCACACACGTGCAGAACACAGGC 854
QY      273 IleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThr 292
DB      853 GTCCCGGAGGCGAGCGCGCTG-----GCGGATTTGTCGCGCTCTTCACGAGGCC 803
QY      293 MetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAsp 312
DB      802 GTGCGGACGTGCTGTCGAAGAG-----CTGTCGCGCGCGCGCGCGCGCG 758
QY      313 LeuLeuProGlnAsnAlaValLeuValAlaSerGlyValAlaSerPheThr 332
DB      757 TTG-----GCCCAAGACAGTTGTGTCTGTCGCGCGCGCTGCGCGCAATCCGCG 707
QY      333 IleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLysProPro 352
DB      706 CTCGCGGACATGTCACGCGCGAGCGGAGGAGGCGGTGAACATGTTCTTCGCCCCG 647
QY      353 ProArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArg 372
DB      646 GTCCGCGTGTGACGAGCAATGCGCATGATTCGCGTGGCGGGGTATGAGCGCTACCGC 587
QY      373 AlaGlyLeu 375
DB      586 CGCGGCGCTG 578

```

RESULT 13  
 US-08-087-797-1  
 ; Sequence 1, Application US/08087797  
 ; Patent No. 5543312  
 ; GENERAL INFORMATION:

```

; APPLICANT: Mellors, Alan
; APPLICANT: Lo, Reggie Y.C.
; APPLICANT: Abdullah, Khalid M.
; TITLE OF INVENTION: Pasteurella Haemolytica
; TITLE OF INVENTION: Glycoprotease
; TITLE OF INVENTION: Gene and the Purified Enzyme
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell, Seltzer, Park & Gibson, P.A.
; STREET: 1211 East Morehead Street,
; CITY: Charlotte
; STATE: No.5543312th Carolina
; COUNTRY: United States
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/087,797
; FILING DATE: 14-JUL-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Layton, Jr., Samuel G.
; REGISTRATION NUMBER: 22807
; REFERENCE/DOCKET NUMBER: 3374-80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 704 377 1561
; TELEFAX: 704 334 2014
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1315 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 141..1115
; US-08-087-797-1
;
; Alignment Scores:
; Pred. No.: 1,19e-50 Length: 1315
; Score: 479.00 Matches: 128
; Percent Similarity: 48.63% Conservative: 50
; Best Local Similarity: 34.97% Mismatches: 157
; Query Match: 22.54% Indels: 31
; Gaps: 8
;
; US-10-649-273-2 (1-414) x US-08-087-797-1 (1-1315)
QY      9 GlyValPhePheLysProSerLysArgLysValTyrGluPheLeuArgSerPheAsp 28
DB      89 GGGCGGTCTTCCGCCCC-----TTGGTTTTCACATTAATTTCATCTTC 133
QY      29 HisProGlyThrLeuPheLeuHisLysIle-ValLeuGlyIleGluThrSerCysAspAs 48
DB      134 TCC-----AACATAGCGAATTTAGGTATGAAACCTTGATGATCA 175
QY      48 PThrAlaAlaValAlaAspGluThrGlyAsnValLeuGlyLysAlaIleHisSerG 68
DB      176 AACCGGTGTGCATTATGATGAGACAAAGGCTTAGGCGCAACAGCTTATAGCCA 235
QY      68 PThrGluValHisLeuLysThrGlyGlyIleValProProAlaIleGlnGlnHisArg 88
DB      236 AATTGATATGACACGCGCATTAAGGTAGCGGTACTGCACTCTTCGACACCAAT 295
QY      88 GlnLysAsnIleGlnArgIleValGlnGluAlaLeuSerAlaSerGlyValSerProSerAs 108
DB      296 CCGTAAGAGCTTGCACATTAATTCAGAGGCTTAAAGAGGCAATTCGCAACCTCGGA 355

```

108 pLeuSerAlaIleAlaThrThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLe 128  
128 userPheSerLeuGlnLeuValGlyGlnLeuLysLysProPheIleProIleHisSme 148  
416 AACCAATGCGCCCTGCTGCTTATGCTTGAATGTTCCGGCATTTGGCGCTTAT 475  
148 tGluAlaHisAlaLeuThrIleArgLeu---ThraSnlLysValGlyPheProPheLeu 167  
476 GGAAGGCGATTACTTCTGCCCCCAATGTTGAAAGAAATGCCCTGAAATTTCTGTTGGC 535  
167 IleuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPhe 187  
536 ATTATTGATTTCAGGAGACACACCAACTGTGTAAGATTGACGGGCTGGGCAATACGA 595  
187 uLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 207  
596 ACTACTCGGGGAAATCAATGATGATCTGCTGGTGAAGCCTTTGACAAACAGGCAACT 655  
207 gLeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyLysAlaIleGluHis 227  
656 ACTCGTTTG---GATTACCT-----GCCGGTGTACGCAATGCAAA 694  
227 sLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 247  
695 ATTAGCCGAATCCGACCCCAATGCTTTAATTCCTCCGTCATGACGACAGACACC 754  
247 sAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLys 267  
755 GGGACTGAGATTTCAGTTCTCCCGTTTAAACCTTTGCGCAATACATTAAGCCAA 814  
267 sGluLysGlnGlyGlyIleGlnLysGlnIleLeuSerSerAlaAlaAspIleAlaI 287  
815 TCTTAATGAATATGTAAGCTCGATGACGAA-----ACCAATGCCATATTGGCCA 865  
287 aThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeu 307  
866 CGCATTCACAAACAGCCGTC-----GTTGATACATATTTAATTA 904  
307 eCysLysGlnArgAspLeuLeuProGlnAsnAlaAlaLeuValAlaSerGlyLys 327  
905 ATGCAAG---CGAGCGTTAGAGCAACCGGCTATTAACCTTAGTATGACAGCGCGCT 961  
327 lAlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCys 347  
962 AAGTCCCAATTAACAATTAACGACAGACCTTGGCGAAATGATGAAAAAATTAAAGCGCA 1021  
347 rLeuLeuCysProProArgLysCysThrAspAsnGlyIleMetIleAlaThrAsnG 367  
1022 AGTATTCTACCTCGCCCAATTTTGACATGACACGGCGCAATGTTCCCTACACTGG 1081  
367 yLleGlnArgLeuArg 372  
1082 CTTTCTCGCTTAAAA 1097

RESULT 14  
US-09-557-884-1/c  
Sequence 1, Application US/09557884  
Patent No. 6506581  
GENERAL INFORMATION:  
APPLICANT: Fleischmann et al.  
TITLE OF INVENTION: The Nucleotide sequence of  
the Haemophilus influenzae Rd Genome, Fragments  
Thereof, and Uses Thereof  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: MD  
COUNTRY: USA

ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3 1/2 inch diskette  
COMPUTER: Dell Pentium  
OPERATING SYSTEM: MS DOS v6.22  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/557,884  
FILING DATE: 25-Apr-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/476,102  
FILING DATE: JUN-5-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Michelle S. Marks  
REGISTRATION NUMBER: 41,971  
REFERENCE/DOCKET NUMBER: PB186P3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-309-8439  
TELEFAX: 301-309-8439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1830121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-557-884-1  
Alignment Scores:  
Pred. No.: 1.21e-44 Length: 1830121  
Score: 475.50 Matches: 120  
Percent Similarity: 50.74% Conservative: 52  
Best Local Similarity: 35.40% Mismatches: 144  
Query Match: 22.38% Indels: 23  
Gaps: 7  
US-10-649-273-2 (1-414) x US-09-557-884-1 (1-1830121)  
QY 39 ValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThrGly 58  
Db 553208 ATCTTAGCATTTGAACCTTCCTGATGAACCGGCGGATTTAGTGAAGAAAA 553149  
QY 59 AsnValLeuGlyGlnAlaIleHisSerGlnThrGlnValHisLeuLysThrGlyLys 78  
Db 553148 GGAATTAATGCTAATCACTTATACATAATTCGCCCTGACAGCATTAATGAGTGTG 553089  
QY 79 ValProProAlaAlaGlnGlnLeuHisArgGlnAsnIleGlnArgIleValGlnGln 98  
Db 553088 GTTCCTGAATTAAGCATCAGCATGATCATATTCGAAAAACAGACCTCTTAATTAAGCGCA 553029  
QY 99 LeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrIleLysPro 118  
Db 553028 TTAGAGGAGACCAATTTAACCGCAGCATTTGATGATGATGATGATGATGATGATG 552969  
QY 119 GlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGln 138  
Db 552968 GGGCTTGTGGCGCATGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 552909  
QY 139 LysLysProPheIleProIleHisHisMetGlnAlaHisAlaLeuThrIleArgLeu 157  
Db 552908 AATGCTCTCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 552849  
QY 158 ThrAsnLysValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeu 177  
Db 552848 GACATTCACCGCACTTCTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 552789  
QY 178 AlaLeuValGlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaPro 197  
Db 552788 GTGCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 552729  
QY 198 GlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysProGluCysSer 217

```

Ddb 552728 GGGCAAGCCTTGATMAAAGACCAAAATTAAGTGGACTA---CATATATCA----- 552681
Oy 218 ThrMetSerGlyGlyValAlaIleGluHisIleuAlaIleGlnIleValAsnArgPheHisIlePhe 237
Ddb 552680 -----GGTGGGCGCGGCTTTCTCGTTTAGCGGAAAGAGTACGCCAAATCGTTTC 552630
Oy 238 AspIleIysProPheLeuHisHisAlaIysAsnCyAspPheSerPheThrGlyLeuGln 257
Ddb 552629 ACATTTCACGTCGATATGACAGATGTCGACAGCGCTTGATTTTAGTCTTTCTGCTTTTAAA 552570
Oy 258 HisValThrAspIleIleIleMetIysGlyIysGluGluGly-----IleGluIys 275
Ddb 552569 ACATTGGCGCAAAATATACAGTTATATCAAGCAATTTAAAACGAGGGCGAATGATATAGCAAA 552510
Oy 276 GlyIleIleLeuSerSerAlaIleAspIleAlaIleThrValGlnHisIleThrMetAlaCys 295
Ddb 552509 -----ACTTAACACAGATATTTGCTTATGCTTTCCAGATGCGGAGTGAT 552465
Oy 296 HisIleValIysArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuLeuPro 315
Ddb 552464 ACTCTTGCC-----ATTAATATGTAAG--CGTGCATTGAAAAGAA 552425
Oy 316 GlnAsnAsnAlaValIleuValAlaSerGlyGlyValAlaSerAsnPheArgIleArgArg 335
Ddb 552428 ACAGCGCTATTAACGTTTAGTGTGATTCGGGAGGGGTAGCGCAATATAAAAACTCCGAGAA 552366
Oy 336 AlaIleGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 355
Ddb 552368 ACGCTTGGCGCACTTAATATGCAAAATTAGGTGGCGAAGCTGTTTATCTCAACCTCATTTT 552309
Oy 356 CysThrAspAsnArgIleIleMetIleAlaIlePheAsnGlyIleGluArgLeuAArgAlaGly 374
Ddb 552308 TGTACAGATTAATAGTGTCGATGATGCTTACACAGGTTTTTTACGTTTAAAACAAGGT 552252

RESULT 15
US-09-643-990A-1/c
Sequence 1, Application US/09643990A
Patent No. 6528289
GENERAL INFORMATION:
APPLICANT: Robert D. Fleischmann
Mark D. Adams
Owen White
Hamilton O. Smith
J. Craig Venter
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS V6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1

```

```

? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 301-610-5790
? TELEFAX: 310-309-8439
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1830121 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Alignment Scores:
Pred. No.: 1,21e-44 Length: 1830121
Score: 475.50 Matches: 120
Percent Similarity: 50.74% Conservative: 52
Best Local Similarity: 35.40% Mismatches: 144
Query Match: 22.38% Indels: 23
DB: 4 Gaps: 7

US-10-649-273-2 (1-414) x US-09-643-990A-1 (1-1830121)
QY 39 ValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValAlaSerGluThrGly 58
Db 553208 ATCTTAGCATTAAGAACTTCCTGTGATGAAGACGGCGCGCATTTATATGAGAAAG 553149
QY 59 AsnValLeuGlyValAlaIleHisSerGlnThrGluValHisLeuYserGlyIle 78
Db 553148 GGATTAATGCTAATCACTTTATACTCAAAATGGCCCTGCATGACAGATTATGAGTG 553089
QY 79 ValProPheAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnAla 98
Db 553088 GTTCTCGAATTAGCATTCACGATATCATATTTCGAGAAACAGACCTCTTATTTAAAGCGCA 553029
QY 99 LeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLeuPro 118
Db 553028 TTAGAGGAAGCAATTTAACGCGACGATTTGATGTGATTGCTTATACGATGGCCCT 552969
QY 119 GlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeu 138
Db 552968 GGGCTGTGTGGGCGCATTTGCTTGGGTGCGGTGTACGATTGACGCTTACCTGCTTGG 552909
QY 139 LysLysProPheIleProIleHisHisMetGluAlaHisAlaLeuThrIleArgLeu--- 157
Db 552908 AATGTTCTCGCATTTGGTGTATCATATGAAAGGCGCATTTACTCCGCCAATGCTGAT 552849
QY 158 ThrAsnLysValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeu 177
Db 552848 GACAAATTCACCGCACCTTCTTTGTGTGCTGTGGTGTGCGGTGCGCACACTCAATTA 552789
QY 178 AlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaPro 197
Db 552788 GTGCGCTTCGAAAGGTGTAGGAAGAAATGAAAGTAAGGAATCTTATATGATGCTGT 552729
QY 198 GlyAspMetLeuAspLysValAlaArgAlaGluSerLeuIleLysHisProGluCysSer 217
Db 552728 GGGCAAGCCTTGTGATTAACACCAAAATTACTTGACATA--GATTATCA----- 552681
QY 218 ThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPhe 237
Db 552680 -----GGTGGCGCGCGCACCTTCTGCTTTAGCGGAGAAAAGTACGGCAATGCTTTC 552620
QY 238 AspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGln 257
Db 552629 ACATTTCCACGCTCAATGACACATGCTGCAGCGCTTGATTTAGTTTCTGTGTTTAAA 552570
QY 258 HisValThrAspLysIleIleMetLysLysGluLysGluGlnGly-----IleGluLys 275
Db 552569 ACATTTGCGCAAAATACGATTATCAAGCAATTTAAAACAGAGGACCACTGATAGACAA 552510
QY 276 GlyGlnIleLeuSerSerAlaAlaSerIleAlaIleAlaThrValGlnHisThrMetAlaCys 295
Db 552509 -----ACTAAAGCATATTTGCTTATGCTTCCAAATACGCGTGTGTGAT 552465

```

```
Qy      296 HisLeuValIysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 315
      552464 ACTCTTGCC-----ATTAAATGTAAAG---CGTCATTCGAAGAA 552429
      316 GlnAsnAsnAlaValIleuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 335
      552428 ACAGGCTATAAACGTTTAGTATGCGGAGGGGTGAGCGCAATAAATAAACTCCGAGAA 552369
Qy      336 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrIleuLeuCysProProArgLeu 355
      552368 AGCCTTGCGCAGCTTAATGCCAAATTAGGTGCGCAAGTGTTTATCTCAACCTCAATTT 552309
      356 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGly 374
      552308 TGTACAGATPATGTGTGCGATGATTGCTTACACAGGTTTTTACGTTTAAACAAGGT 552252
Db
```

Search completed: June 17, 2005, 02:00:31  
Job time : 1592.65 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: June 16, 2005, 17:01:19 ; Search time 3547.8 Seconds  
(without alignments)  
4441.797 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125  
Sequence: 1 MLITFTAGVFPKPSKRYV.....DISKEVGEASIKVPOLMEI 414

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=xlp  
-Q=/cgn2.1/USPTO.spool.p/US10649273/runat.15062005.111416.6043/app.query.fasta\_1.1429  
-DB=EST -QPMT=fastaap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bites -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTPMT=pro -NORH=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10649273.@CEN.1.1.8076.@runat.15062005.111416.6043 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST: \*  
1: gb\_est1: \*  
2: gb\_est2: \*  
3: gb\_hic: \*  
4: gb\_est3: \*  
5: gb\_est4: \*  
6: gb\_est5: \*  
7: gb\_est6: \*  
8: gb\_gsa1: \*  
9: gb\_gsa2: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score  | Query Match | Length | DB | ID       | Description        |
|------------|--------|-------------|--------|----|----------|--------------------|
| 1          | 1827   | 86.0        | 2284   | 3  | AK045669 | AK045669 Mus muscu |
| 2          | 1819   | 85.6        | 1622   | 3  | AK011265 | AK011265 Mus muscu |
| 3          | 1634   | 76.9        | 1609   | 3  | BC030671 | BC030671 Mus muscu |
| 4          | 1447   | 68.1        | 870    | 5  | BQ423651 | BQ423651 AGENCOURT |
| 5          | 1353.5 | 63.7        | 852    | 5  | BX391919 | BX391919 AGENCOURT |
| 6          | 1279   | 50.2        | 322    | 5  | BQ561026 | BQ561026 AGENCOURT |
| 7          | 1173   | 55.2        | 701    | 2  | BE740611 | BE740611 601595739 |
| 8          | 1158   | 54.5        | 749    | 2  | CK982692 | CK982692 4115331 B |
| 9          | 1144   | 53.8        | 800    | 7  | CO738006 | CO738006 SLIH03c19 |

|    |        |      |      |   |          |                     |
|----|--------|------|------|---|----------|---------------------|
| 10 | 1132   | 53.3 | 1109 | 5 | BM907988 | BM907988 AGENCOURT  |
| 11 | 1111   | 52.3 | 765  | 5 | CF114247 | CF114247 Shultzomi  |
| 12 | 1097   | 51.6 | 640  | 5 | BQ636028 | BQ636028 hd03d11.y  |
| 13 | 1045   | 49.2 | 730  | 7 | CN823245 | CN823245 Oa.sp1bn   |
| 14 | 1038.5 | 48.9 | 661  | 7 | CK833139 | CK833139 4057170 B  |
| 15 | 1018.5 | 47.9 | 879  | 5 | BQ433135 | BQ433135 AGENCOURT  |
| 16 | 1017   | 47.9 | 1173 | 6 | CD508917 | CD508917 CDA93-E05  |
| 17 | 997    | 46.9 | 723  | 7 | CK942291 | CK942291 4065791 B  |
| 18 | 978    | 46.0 | 579  | 7 | CK819035 | CK819035 1f06d05.y  |
| 19 | 966    | 45.5 | 822  | 7 | CF257246 | CF257246 pha008.g0  |
| 20 | 963    | 45.3 | 577  | 7 | CV028547 | CV028547 7089.Full1 |
| 21 | 954.5  | 44.9 | 808  | 7 | CK792978 | CK792978 AGENCOURT  |
| 22 | 954    | 44.9 | 833  | 5 | BU159877 | BU159877 AGENCOURT  |
| 23 | 951.5  | 44.8 | 728  | 4 | BG436856 | BG436856 60248845   |
| 24 | 935.5  | 44.0 | 879  | 5 | BU256052 | BU256052 603746167  |
| 25 | 935    | 44.0 | 871  | 2 | BF308169 | BF308169 601887411  |
| 26 | 925    | 43.5 | 658  | 7 | CF362328 | CF362328 829596.MA  |
| 27 | 923    | 43.4 | 811  | 7 | CN232675 | CN232675 MLB051B07  |
| 28 | 920    | 43.3 | 859  | 5 | BU292595 | BU292595 603604785  |
| 29 | 916    | 43.1 | 831  | 7 | CK455183 | CK455183 915735.MA  |
| 30 | 915    | 43.1 | 812  | 5 | BU246489 | BU246489 603784202  |
| 31 | 912    | 42.9 | 878  | 7 | CK456228 | CK456228 916887.MA  |
| 32 | 911    | 42.9 | 661  | 7 | CF787040 | CF787040 854903.MA  |
| 33 | 907    | 42.7 | 717  | 7 | CK459683 | CK459683 929895.MA  |
| 34 | 906    | 42.6 | 597  | 6 | CB272391 | CB272391 ma157g12.  |
| 35 | 902    | 42.4 | 533  | 4 | BM126453 | BM126453 1f06d05.y  |
| 36 | 893    | 42.0 | 574  | 5 | BM130908 | BM130908 1f89c10.y  |
| 37 | 893    | 42.0 | 687  | 7 | CF360919 | CF360919 827217.MA  |
| 38 | 879    | 41.4 | 737  | 7 | CV073257 | CV073257 AGENCOURT  |
| 39 | 867    | 40.8 | 616  | 6 | CB438468 | CB438468 687077.MA  |
| 40 | 857.5  | 40.4 | 789  | 5 | BU242187 | BU242187 603781023  |
| 41 | 838    | 39.4 | 637  | 7 | CK941819 | CK941819 4065407.B  |
| 42 | 831    | 39.1 | 853  | 7 | CR444993 | CR444993 CR444993   |
| 43 | 825.5  | 38.6 | 806  | 4 | BG333362 | BG333362 602431206  |
| 44 | 819.5  | 38.6 | 792  | 5 | BU403563 | BU403563 604138456  |
| 45 | 819    | 38.5 | 668  | 5 | BU285782 | BU285782 604166893  |

## ALIGNMENTS

RESULT 1  
AK045669  
LOCUS  
DEFINITION  
AK045669 2284 bp mRNA linear HTC 03-APR-2004  
Mus musculus adult male corpora quadrigenima cDNA, RIKEN  
full-length enriched library, clone.B230219017 product: similar to  
PUTATIVE STATOGLYCOPROTEIN TYPE 2 [Homo sapiens], full insert  
Sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AK045669  
AK045669.1 GI:26337528  
HTC; CAP trapper.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
10349636  
2  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
JOURNAL  
MEDLINE  
PUBMED  
20499374  
11042159  
3  
Shimada, K., Itoh, M., Atzawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitanai, T., Tashiro, H., Itoh, M.,  
Sun, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
10349636  
2  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
JOURNAL  
MEDLINE  
PUBMED  
20499374  
11042159  
3  
Shimada, K., Itoh, M., Atzawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitanai, T., Tashiro, H., Itoh, M.,  
Sun, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

|           |   |
|-----------|---|
| TITLE     | Pujiwaka, S., Inoue, K., Togawa, Y., Iwawa, M., Ohara, E., Matshiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kiz, A. and Hayashizaki, Y.  |
| JOURNAL   | RIKEN Integrated sequence analysis (RISA) system--384-format  |
| MEDLINE   | sequencing pipeline with 384 multicapillary sequencer   |
| PUBMED    | Genome Res. 10 (11), 1757-1771 (2000)   |
| REFERENCE | 20530913  |
| AUTHORS   | 11076661  |
| TITLE     | 4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  |
| JOURNAL   | Functional annotation of a full-length mouse cDNA collection  |
| REFERENCE | Nature 409, 685-690 (2001)  |
| AUTHORS   | 5   |
| TITLE     | The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  |
| JOURNAL   | Analysis of mouse transcriptome based on functional annotation  |
| REFERENCE | of 60,770 full-length cDNAs   |
| AUTHORS   | Nature 420, 563-573 (2002)  |
| TITLE     | 6 (bases 1 to 2284)   |
| JOURNAL   | Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hizumoto, K., Hirooka, T., Hirozane, T., Horii, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.   |
| COMMENT   | Direct Submission<br>Submitted (16-Jul-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  |
| FEATURES  | cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.<br>Please visit our web site for further details.<br>URL: http://genome.gsc.riken.jp/<br>URL: http://fantom.gsc.riken.jp/.<br>Location/Qualifiers  |
| SOURCE    | 1. 2284<br>/organism="Mus musculus"<br>/mol_type="mRNA"<br>/strain="CS7BL/6J"<br>/db_xref="PANTOM,DB:B230219017"<br>/db_xref="taxon:10090"<br>/clone="B230219017"<br>/sex="male"<br>/tissue_type="corpora quadrigemina"<br>/clone_lib="RIKEN full-length enriched mouse cDNA library"<br>/dev_stage="adult"<br>221. 1465<br>/note="unnamed protein product; putative<br>similar to PUTATIVE SIALICINOPROTEASE TYPE 2 (Homo sapiens) (SPTR Q9H4B0, evidence: FASTV, 80%ID, 100%length, match=1242)"<br>/codon_start=1<br>/protein_id="BAC32450.1"<br>/db_xref="gi:26337529"<br>/translation="MLMRLPRAGAIIPKPKSKYVGLRFSVHPTLSLGLVIGIET<br>SCDDTAAVDEGNGVIGEALHSOTQVHLKTGIVPPAOLQHEKRIORIYEETLSAQ<br>RTPSDLSAIVTITKPGALSLGUGLSSTLOLVNOFKPPIPHHMEHATITLTKR<br>VEPFLVLSGGHCLALVGVSDFFILGSLDIPEDMDLQAKRRSLTKHPECST<br>MSGKALVQLAKDGNRFHFTIPVQNAKNDVFSFTQHTITDLITRKEKEBIEKGG |

|   |     |   |               |  |  |
|---|-----|---|---------------|--|--|
|   |     | polyA_signal  |               | OIISSADVIAAAVQHTATCAKRTTHAIIIFCQKKNLSPPANAALVWSGVAASMLYR |  |
|   |     | 2267..2267  |               | KALEIVANQCTLLCPRPRLCTDNGMIAMNGIERLRAGLGVLAHVEDIHYERKCPLE |  |
|   |     | polyA_site  |               | GVDSRSRYAAEAIIVPRLLKAL"                                  |  |
|   |     | /note="putative"  |               |  |  |
|   |     | 2284  |               |  |  |
|   |     | /note="putative"  |               |  |  |
| ORIGIN                                      |     |   |               |  |  |
| Alignment Scores:                           |     |   |               |  |  |
| Pred. No.:                                  |     | 5.12e-194   | Length:       | 2284   |  |
| Score:                                      |     | 1827.00   | Matches:      | 352  |  |
| Percent Similarity:                         |     | 90.58%  | Conservative: | 23   |  |
| Best Local Similarity:                      |     | 85.02%  | Mismatches:   | 39   |  |
| Query Match:                                |     | 85.98%  | Indels:       | 0  |  |
| DB:   |     | 3   | Gaps:         | 0  |  |
| US-10-649-273-2 (1-414) x AK045669 (1-2284) |     |   |               |  |  |
| QY  | 1   | MetLeuIleLeuThrIlySThrAlaGlyValPheNheIysProSerIysArgLysValTyr     | 20            |  |  |
| DB  | 221 | ATGCTAAAGTTTAAAGAACAGACAGAGAGACTATTCCTCCAGCCCCCAAAAGTAAGTTTAT     | 280           |  |  |
| QY  | 21  | GlupPheLeuArgSerPheAsnPhaHisProGlyThrLeuPheLeuHisIlysIleValLeu    | 40            |  |  |
| DB  | 281 | GGATTTTAAAGAAGATTTAAGTTCCTACCTCCAGAACTCTCTTGGTCATTAACGTGCTCTG     | 340           |  |  |
| QY  | 41  | GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal         | 60            |  |  |
| DB  | 341 | GGAAATTGAAACACAGCTGTGATGACACAGAGACGGCTGTGTGTGATGAATACTGGAAATGTG   | 400           |  |  |
| QY  | 61  | LeuGlyGluAlaIleHisSerGlnThrGluValHisIleuLysIlyThrGlyGlyTleValPro  | 80            |  |  |
| DB  | 401 | CTGGGGGAGACAGCTGCACCTCCCAACTCAGGTTTCATCGAAAACAGGGGGAATGTTTCT      | 460           |  |  |
| QY  | 81  | ProAlaIaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer       | 100           |  |  |
| DB  | 461 | CGAGTAGCTCAACAACTTCACAGAGAAATATTCACAGAAATGTAAGAAACCTTTCT          | 520           |  |  |
| QY  | 101 | AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu      | 120           |  |  |
| DB  | 521 | GCGCTAGAAATCACCCCAAGGAGATCTCTACAAATTGCACTTACCATCAAAACCGGAGCTG     | 580           |  |  |
| QY  | 121 | AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys         | 140           |  |  |
| DB  | 581 | GCCCTAAGCTTGGAGATGGCTTATCCTTTAGCTTACAGCTAGTAATAACAGTTTAAAAAG      | 640           |  |  |
| QY  | 141 | ProPheIleProIleHisIleMetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLys      | 160           |  |  |
| DB  | 641 | CCATTTCACCCCGATTCAATCACATGAGAGCTACAGCATGACTATTATTAAGGCTCACCAATAA  | 700           |  |  |
| QY  | 161 | ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal      | 180           |  |  |
| DB  | 701 | GTAGAATTTCTCTTTTTRAGTCTTTTATTTCTGCGGCTACATGCGCTGTGGCATTAATC       | 760           |  |  |
| QY  | 181 | GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet      | 200           |  |  |
| DB  | 761 | CAAGGTGTTTCCGATTTCTCTCTCTCTGGGAAGCTTTTGACATATGACACAGGCCAGCATAG    | 820           |  |  |
| QY  | 201 | LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluLysSerThrMetSer      | 220           |  |  |
| DB  | 821 | CTTGACAAAGGTGGCAAGAAAGCTTTCTTTAATCAAAACATCCAGAAATGTTCTTACAAATGAGT | 880           |  |  |
| QY  | 221 | GlyGlyLysAlaIleGluHisIleuAlaLysGlnGlyAsnArgPheHisPheAspIleLys     | 240           |  |  |
| DB  | 881 | GGTGGAAAGCTTAAGAACAGTGGCCCAAGAGGAATTAATTCATTTTAATCAT              | 940           |  |  |
| QY  | 241 | ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr      | 260           |  |  |
| DB  | 941 | CAACTATGACGAATGCTTAAGATATGCAATTTTCTTTTCAACGGAGCTTCAACATATTAAT     | 1000          |  |  |
| QY  | 261 | AspLysIleIleMetLysGlyGluLysGluGluGlyIleGlyLysGlyGlnIleLeuSer      | 280           |  |  |

[illegible]

TITLE  
JOURNAL

REFERENCE  
AUTHORS

JOURNAL  
AUTHORS

TITLE  
JOURNAL

COMMENT

FEATURES  
SOURCE

CDS

polyA\_signal  
polyA\_site

1622

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

6 (bases 1 to 1622)

Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukushima,Y., Furumoto,M., Hanagaki,T., Hara,A., Hayatsu,N., Hizamoto,K., Hisabata,T., Horii,F., Imotani,K., Ishii,Y., Itchihashi,M., Izawa,K., Kanekura,T., Kato,H., Kawai,Y., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Saeki,D., Shibata,K., Shidatta,Y., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

Direct Submission

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute: 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

Please visit our web site (<http://genome.gsc.riken.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedica Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAAGACATCCAGACTCTTTTTCCTTTTTTTTAA 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCTCGATTAAATTAATTAATGCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.

Host: SOLR.

Location/Qualifiers

1..1622

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="PANTOM DB:2610001M19"

/db\_xref="taxon:10090"

/clone="2610001M19"

/tissue\_type="whole body"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="10 days embryo"

207..1451

/note="unnamed protein product, putative sapiensl (SPTR|QH4B0, evidence: FASTY, 80%ID, 100%length, match=1242)"

/codon\_start=1

/protein\_id="BAB27506.1"

/db\_xref="gi:12847276"

/translation="MLMKRRKAGALIPRPSSKYGVFLRSVHRTLSCHKLVIQGITSCSDTGAAVDENGVNLEGAASQTVALKTGGVLPVPAQDLSHEINQRIVETLTKCRITPTSDLAIAVTIKPGIALSLGVGLFSFLOLVNQPFKPIPHMEAHALTLRLTKAVERPFPILILSGCHLADLVGSDFLLGLSLDIAPDKMDIVARSLSLIKPECGTMSGKAIBOLAKGNRHFTNPMPONAKNDPSFGIOHLITDLITHKEEBEIIKKMOISSSADIAPAAVOHATRHCLAKTHRALILECKRNLSPPANAVIVSGVAASLYLRKALTIVANAQVLTLCPEPLCTMGNGINIANVGSRUNCLCVLDHYEDINIVPKCTLGVDSREVAEAIVPRILKAL"

1605..1610

/note="putative"

1622



ORIGIN /note="putative"

## Alignment Scores:

Pred. No.: 2 38e-193 length: 1622  
Score: 1819.00 Matches: 351  
Percent Similarity: 90.34% Conservative: 23  
Best Local Similarity: 84.78% Mismatches: 40  
Query Match: 85.60% Indels: 0  
Gaps: 0

US-10-649-273-2 (1-414) x AK011265 (1-1622)

```

Qy 1 MetLeuIleLeuThrIsthrAlaGlyValaPhePheLysProSerLysArgLysValIle 20
Db 207 ATGCTAATGTTAAGAAGCAGACGAGCATTCCTCCAGGCCCAAGAGTAAAGTTAT 266
Qy 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40
Db 267 GGATTTTAAAGAAAGATTAGTGTTCATCCAGAACTCTCTTGTCTAATAACTGCTCTG 326
Qy 41 GlyIleGluThrSerGysAspAspThrAlaAlaValaValaSerGluThrGlyValaVal 60
Db 327 GGAATTGAAACCAAGCTGTGTATGACACAGAGCGGCTGTGTGATGAAACTGGGAATGTG 386
Qy 61 LeuGlyValaAlaIleLeuSerGlnThrGluValaHisIleuLysThrGlyIleValPro 80
Db 387 CTGGGGGAAACAGCTGACATCCCAAACTCAGGTTCTATGAAAACAGGTGGATTGTTCT 446
Qy 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValaGlnGluValaLeuSer 100
Db 447 CCAGTAGCTCAACAACTTCACAGAGAAATATTCACAGAAATAGTAGAAGAACTTTCT 506
Qy 101 AlaSerGlyValaSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
Db 507 GCCTTAGATATCAACCCCAAGCGATCTCTGACAAATTCACATCCATCAACCGGAGCTG 566
Qy 121 AlaLeuSerLeuGlyValaGlyLeuSerPheSerLeuGlnLeuValaGlyIleLeuLysLys 140
Db 567 GCCCTTAGCTGGAGATTTGGCTTATCTTACCTTAGCTAGTAATCATGTTTAAAG 626
Qy 141 ProPheIleProIleHisIleHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
Db 627 CCATTATCCCATTCATCATCAATGAGAGGCTCAGCACTGATATTAGGCTCACCAATAAA 686
Qy 161 ValGluPheProPheLeuValaLeuLeuIleSerGlyGlyHisCysLeuLeuValaLeuVal 180
Db 687 GTAGAAATTCCTTTTATGTTTATGATTCTGCGCGTCACTGCTGTTGCAATTAGTC 746
Qy 181 GlnGlyValaSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
Db 747 CAAGGTTTCCATTTCTGCTCTCTGGAGAGCTTTGACATAGACACAGGAGCATG 806
Qy 201 LeuAspLysValaAlaArgArgLeuSerLeuIleLysIleProGlyCysSerThrMetSer 220
Db 807 CTTGACAAAGTGCACAGACATTTCTTAAATCAACATCCAGAAATTTCTTACATAGT 866
Qy 221 GlyIlyLysAlaAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
Db 867 GGTGGAAGAGCTATAGAACAGTTGGCCAAAGACGAAATAGATTCATTATCATCAAT 926
Qy 241 ProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGlyLeuGlnHisValThr 260
Db 927 CCACCTATCAGAAATCTAAGAAATTTGGAATTTCTTTCACGGGACTTCAACATTAAT 986
Qy 261 AspLysIleIleLysMetLysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSer 280
Db 987 GATTAAGCTAATAACACACAGAGAAAGAAAGAGCACTTAGAAGAGCAAAATCTCTCA 1046
Qy 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300
Db 1047 TAGGCTGAGAGACTTCTGCTGCGGTACAGCATGCAACAGCGTCCACTTGGGAAAGAA 1106

```

```

Qy 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320
Db 1107 ACACATGCGCCTATCTCTGTTTGCAGACAGAAAAATTTGCTCTCCAGCTAACGCACTTA 1166
Qy 321 LeuValaAlaSerGlyValaAlaSerAspPheThrIleArgArgAlaLeuGlnIleLeu 340
Db 1167 TTAGTTTATCTTGAAGGTTGTGCAAGTAACTTGTACATCCAGAAAGCAATGGAATTTCTC 1226
Qy 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProAlaGlyLeuCysThrAspAsnGly 360
Db 1227 GCAAATGCAACGACGTCACGCTTTGTGTCTCCTCCAAAGCTGTGACATGACATAGGC 1286
Qy 361 IleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
Db 1287 ATCATGATTCATGAGTAATGGAATTTGAAGATTACGTGCTGGCTGGCGGCTTTATCATAT 1346
Qy 381 IleGluGlyIleArgGlyTrpGluProLysCysProLeuGlyValaAspIleSerLysGluVal 400
Db 1347 GTACAGACACATCCATATGAAACCAAAATGTCCTTGTGAGTAGACATATCCAGAGAAATT 1406
Qy 401 GlyIleAlaSerIleLysValaProGlnLeuLysMetGluIle 414
Db 1407 GCAGAGCTGCTATTAAGATACCGCATTAATAATGGCATT 1448

```

RESULT 3  
BC030671  
LOCUS  
DEFINITION Mus musculus O-sialoglycoprotein endopeptidase-like 1, mRNA (cDNA clone IMAGE:122618), containing frame-shift errors.  
ACCESSION BC030671  
VERSION BC030671.1 GI:21040459  
KEYWORDS HTC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E., Brownstein M.J., Umed T.B., Tosiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J., Bosak S.A., McKean P.J., Mckernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huilyk S.W., Vallalao D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smolins D.E., Scherch A., Schein J.E., Jones S.J. and Merrin M.A.  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL MEDLINE  
PUBMED 22388257  
12477932  
REFERENCE  
2 (bases 1 to 1609)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (20-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK  
COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Marcello Bento Soares, Ph.D.  
cDNA Library Preparation: M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
 Contact: anand@systemsbiology.org  
 Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRK Plate: 66 Row: e Column: 10  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21312463  
 This clone has the following problem: frame shifted.

## FEATURES

## Source

1. 1609  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /db\_xref="IMAGE:1226118"  
 /clone\_type="Thymus gland, mouse"  
 /issue\_type="Soares\_thymus\_2NMT"  
 /lab\_host="DH10B"  
 /note="Vector: pTT3-Pac"

## ORIGIN

## Alignment Scores:

Pred. No.: 1 64e-172 Length: 1609  
 Score: 1634.00 Matches: 327  
 Percent Similarity: 83.86% Conservative: 21  
 Best Local Similarity: 78.80% Mismatches: 40  
 Query Match: 76.89% Indels: 27  
 Gaps: 2

US-10-649-273-2 (1-414) x BC030671 (1-1609)

QY 1 MetLeuIleuThrIleuValPhePheLeuSerProSerIleuValIleuValTyr 20  
 Db 229 ATCTATATGTTAAGAGAGACAGAGAGATTCCTCCAAAGCCCAAGAGATTAAT 288  
 QY 21 GluPheLeuArgSerPheAenPheHisProGlyThrLeuPheLeuHisIleValIleu 40  
 Db 289 GGAATTTTAAAGAGATTAGTGTCTCATCCAGAACTCTCTGTCTAAATGCTCTG 348  
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60  
 Db 349 GGAATTGAAACCGAGCTGTGATGACACAGAGCGGCTGTGGATGAAATCGGAGATGTG 408  
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisIleu-LysThrGlyGlyIleValIle 80  
 Db 409 CTGGGGGAGACAGCTGCACTCCCAACTCAGGTTCACTGAAAGT-----GGGATTTTCC 462  
 QY 80 oProAlaAlaGlnIleuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSe 100  
 Db 463 TCAGATGAGCTCAACAACTTCAGAGAGAAATATTCAAGATATGAGAGAACTCTTC 522  
 QY 100 rAlaSerGlyValSerProSerSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLe 120  
 Db 523 TGCCTGAGATCAACCCCAAGCAGTCTCTGCAATTGCACTACATCAAAACCGGAGCT 582  
 QY 120 uAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnIleuValGlyGlnIleuVal 140  
 Db 583 GGGCCCTAAGCTGGAGAGTGTCTTATCTTTAGCTTACAGTAAATCAAGTTTAAAAA 642  
 QY 140 sProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnIle 160  
 Db 643 GCCATTCATCCGATTCATCAATGAGGCTCAGCCATGCTATTTAGGCTCACCATAA 702  
 QY 160 sValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
 Db 703 AGTGAATTCCTTTTATGTTCTTTTATTTCTTGGGCTACATGCGCTTGGCATTTAGT 762  
 QY 180 lGlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMe 200

Db 763 CCAAGGCTTTCCGATTTCCGCTCTCTGGAGAGCTTTGGACATAGACACAGGAGCAT 822  
 QY 200 tLeuAspIleValAlaArgArgLeuSerLeuIleLysHisProGlyCysSerThrMetSe 220  
 Db 823 GCTTGACAGAGTGGCAAGAAACATTTCTTTATCAATCAATCATCCAGATATGTTCAATAG 882  
 QY 220 rGlyGlyValAlaIleGluHisLeuAlaValGlnGlyAsnArgPheHisPheAspIleLe 240  
 Db 883 TGGTGAAGAAAGCTTATAGAACAGTTGGCCAAAGACGGAATATGATTCATTTTACTATCA 942  
 QY 240 sProLeuLeuHisIleAlaValAsnGlyAspPheSerPheThrGlyLeuGlnHisValThr 260  
 Db 943 TCCACTATGCAAGATGCTAAGATGAGATTTTCTTTCACGGACCTTCAACATATTC 1002  
 QY 260 rAspIleIleIleMetLysGlyGlyValGlyGlyGlyIleGluLysGlyGlnIleLeuSe 280  
 Db 1003 TGATTAAGCTATTAACACACAGAGAAAGAAAGAGCATTTGAGAGGGGCAATCTGTCTC 1062  
 QY 280 rSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValIleArg 300  
 Db 1063 ATCAGCTGACAGACATTTGCTGCTGGCGGTACAGCATGCAACAGCGTCCACCTTGCAGAAAG 1122  
 QY 300 gThrHisArgAlaIleLeuPheCysGlyGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320  
 Db 1123 AACACATCGCCCTATTTCTGTTTGCAGACAGAAATATTTCTCTCCAGCTTAACGCAGT 1182  
 QY 320 lLeuValAlaSerGlyGlyValAlaSerAspPheThrIleArgArgAlaLeuGluIleLe 340  
 Db 1183 ATTGATGTATCTTGAAGGTGTGTCAGAGTACTTGTGACATCCGAAAGCATTTGAAATTC 1242  
 QY 340 uThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnG 360  
 Db 1243 CGCAATGCAACGAGCTGACAGTGTGTTGTGTCCACTCCAAAGCTGTGACTGACATATCG 1302  
 QY 360 yIleMetIleAlaThrPheGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAs 380  
 Db 1303 CATCATGATTCGA-----TGATTCCTTGTGAGAGATTCATTCACAGAAAGT 1315  
 QY 380 pIleGluGlyIleArgTyrgIuProLysCysProLeuGlyValAspIleSerLysGluVal 400  
 Db 1316 -----TGATTCCTTGTGAGAGATTCATTCACAGAAAGT 1350  
 QY 400 lGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414  
 Db 1351 TGCAGAGCTGCCATTAAGTACCCGATTTAAATGGCATT 1393

RESULT 4  
 B0423651 870 bp mRNA linear EST 23-MAY-2002  
 LOCUS AGENCOURT\_7790948 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:6065828  
 DEFINITION 5', mRNA sequence.  
 ACCESSION B0423651  
 VERSION B0423651.1 GI:21118966  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 870)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 AUTHORS Contact: Robert Strausberg, Ph.D.  
 TISSUE Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: ATCC/DCTD/BTP  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: LAM13342 row: 1 column: 21  
 High quality sequence stop: 710.

# FEATURES source

Location/Qualifiers

1..870

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:6065828"

/cissue\_type="melanotic melanoma"

/lab\_host="DHI0B (phage-resistant)"

/note="Organ: Skin; Vector: pCMV-SPORT6; Site: 1; NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 kb. Library constructed by Life Technologies."

## ORIGIN

### Alignment Scores:

Pred. No.: 7,29e-152 Length: 870

Score: 1447.00 Matches: 285

Percent Similarity: 98.62% Conservative: 1

Best Local Similarity: 98.28% Mismatches: 2

Query Match: 68.09% Indels: 2

DB: Gaps: 0

US-10-649-273-2 (1-414) x BQ423651 (1-870)

111 AAlaAlaThrThrIleLeuProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPhe 130

1 GCATTGCAACTCCATTAACACGAGCTGTTAAGCTGGAGTGGCTTATCATTT 60

131 SerLeuGlnLeuValGlyLeuValLeuPheProPheIleProIleHisSmetGluAla 150

61 AGCTTACAGTGTAGAGACAGTTAAAGCCATTCATTCATTCATCATATGAGGCT 120

151 HisAlaLeuThrIleLeuGlnLeuValGlyLeuPheProPheLeuValLeuIle 170

121 CAGCACTTACATTAATGAGTTGACCAATTAAGTAGAATTCCTTTTAACTTTTAAAT 180

171 SerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGly 190

181 TCTGGAGTCACTGCTGCTGTTGGCATTTAGTTCAAGAGCTTCAAGTTTCTGCTTGA 240

191 LysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeu 210

241 AAGTCTTTGGACATAGACAGGAGGATGATCTTGACAAAGGTGGCAAGACCTTTCTTTA 300

211 IleLeuHisProGlyCysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLys 230

301 ATTAACATCCAGAGCTCTCCACCATGAGTGGGAAAGCCATTAAGCAATTTGGCCAAA 360

231 GlnGlyAsnArgPheHisPheAspIleLeuProLeuHisHisAlaLysAsnCysAsp 250

361 CAAAGAAATGATTTCAATTTGACATCAACCTCCCTTGACATCAATGCTTAATAATTTGAT 420

251 PheSerPheThrCysLeuGlnHisValThrAspLysIleIleMetLysLysGluLysGlu 270

421 TTTTCTTTTACTGACTCAACGCTTACTGATTAATAATTAAGAAAAGAAAAGAG 480

271 GlnGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGln 290

481 GAAGGATTTGAGAAAGGCAATCTGCTTTCAGACACAGACATTCGTCACAGTACAG 540

291 HisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysValGln 310

541 CACACAAATGCAATGCTCTTGTGAAGAAAGAACATCGGGCTAATCTGTTTGTAAAGAG 600

311 ArgAspLeuLeuProGlnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsn 330

601 AGGAGCTTGTACTCAATAATATGCAATCTGTTGATCTGTTGCTGCGCAAGTAC 660

331 PheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCys 350

661 TTCTATATCGCAGAGCTGGAATTTTAACAAACCAACACAGTGCACATTTGGTGTGT 720

351 ProProProArgLeuCysThrAspArgGlyIleMetIleAlaIleArgGlyIleGluArg 370

721 CCTCTCTCCAGACTATCATCATTAATGCAATTAATGATGATGATGATTAATGAAGA 780

371 LeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgGlyIleProLysCys 390

781 CTAGCTCTGCTGGCATTTTACATGACATTAAGAGCATCCCTATGAAACCAATATG 840

390 sPro-LeuGlyValAspIleSerLys 398

841 TCTCTCTGAGTACATATCAAAA 866

## RESULT 5

BX391919 852 bp mRNA linear EST 28-APR-2004

LOCUS BX391919 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens

DEFINITION CDNA clone CS0DK001YE02 3-PRIME, mRNA sequence.

ACCESSION BX391919

VERSION BX391919.2 GI:46846154

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Bukaryaota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 852)

AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT On May 13, 2003 this sequence version replaced gi:30611736.

Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and Bclor V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 1240.r

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?cs=CS0BA1035ZE07\_CS03317\_1&c=1240.r

# FEATURES source

Location/Qualifiers

1..852

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0DK001YE02"

/cell\_type="HELA CELLS COT 25-NORMALIZED"

/clone\_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and Bclor V sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Alignment Scores:

Pred. No.: 2.41e-141 Length: 852

Score: 1353.50 Matches: 274

Percent Similarity: 96.14% Conservative: 0

Best Local Similarity: 96.14% Mismatches: 8

Query Match: 63.69% Indels: 4

DB: Gaps: 2

US-10-649-273-2 (1-414) x BX391919 (1-852)

72 HisLeuLys---ThrGlyGlyIleValProProAlaIleGlnGlnLeuHisArgGluAsn 90

850 CATTAAACACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 798

91 IleGlnArgIleValGlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeuSer 110

```

Db      797 ATTCACGATAGTACAGAAAGCTCTTCTGCGAGTGNAGTCTCTCCAGTACCTCTCA 738
Qy      111 AAllelaThrThrlleYpsProGlyLeuAlaIleuSerleuGlyValGlyLeuSerPhe 130
Db      737 GCATTTGCANCAACCAATTAACCAAGNACTGCTTTTAAGCTGGAGGGGCTTATCATTT 678
Qy      131 SerleuGlnLeuValGlyLeuValLeuValSerPheIleProIleHisIleMetGluAla 150
Db      677 ACCTTACAGCTGGTAGAGACAGTGTAAAAAAGCCATTCTCCATTCAATCATATGAGGCT 618
Qy      151 HsAlaIleuThrlleArgLeuThraenlyValGlnPheProPheLeuValIleuLeuIle 170
Db      617 CATGCACTTACTATTAATTTGTTGACCAATTAAGTAAGTAATTTCTTTTATTTGTTTATTT 558
Qy      171 SerGlyGlyHisCysleuLeuAlaIleuValGlnGlyValSerAspPheLeuLeuGly 190
Db      557 TCTGGAGGCTACTGCTGCTGTTGGCATTTGTTCAAGAGATTTCAGATTTTCTGCTTCTTGA 498
Qy      191 LysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeu 210
Db      497 AGCTCTTTGGACATACACACAGGTGACATCTTGGACAGAGTGGCAAGAGCTTTCTTTTA 438
Qy      211 IleYHisIleProGlyCysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLys 230
Db      437 ATTAACATCTCAAGAGTCTCCACCATGAGTGTGGGAAGCCATAGAACATTTGGCCAAA 378
Qy      231 GlnGlyAsnArgPheHisPheAspIleLysProProIleHisIleAlaLysAsnCysAsp 250
Db      377 CAGGAAATAGATTTCAATTTGACATCAACCTCTTGCATCATGCTTAATAAATTTGAT 318
Qy      251 PheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLysGlyLysGlu 270
Db      317 TTTCTTTTACTGACTTCAACAGCTTACTGATTAATTAATTAAGAAAAAGAAAAAGAG 258
Qy      271 GlnGlyIleGlyLysGlyGlnIleLeuSerSerAlaIleAspIleAlaIleThrValGln 290
Db      257 GAAGGATTTGAGAAAGGGCAATCTCTCTTCCAGCAGACAGATTTGGCCAGTACAG 198
Qy      291 HisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGln 310
Db      197 CACACATGCGCATGCTCATCTTGTGMAAAGAACACATCGGGCTATCTGTTTGTAGAGAG 138
Qy      311 ArgAspLeuLeuProGlnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsn 330
Db      137 AGGAGCTTTGTACTCTCAAAATTAATGACAGTCTGTCATCTGAGTGGCAAGTAC 78
Qy      331 PheTyrlleArgArgAlaLeuGlnIleLeuThraenAlaIleArgIleCysThrLeuLeu 350
Db      77 TTCTATATCCGACAGCTCTGGAATTTTAACAAACCAACACAGTGCATTTGTGTGT 18
Qy      351 ProProProArgLeu 355
Db      17 CTCCTCCCGACACTA 3

```

RESULT 6  
 B0961028 LOCUS 922 bp mRNA linear EST 21-AUG-2002  
 DEFINITION AGENCOURT 8863711 NIH\_MGC\_18 Homo sapiens cDNA clone IMAGE:6423902  
 5' mRNA sequence.  
 ACCESSION B0961028  
 VERSION B0961028.1 GI:22376506  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 922)  
 AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: DCTD/DTP/Gazdar

CDNA Library Preparation: Rubin Laboratory  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
 Plate: LUCM2603 row: d column: 15  
 High quality sequence stop: 584.  
 Location/Qualifiers  
 1..922  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6423902"  
 /cuisine\_type="large cell carcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_id="NIH\_MGC\_18"  
 /note="Organ: lung; Vector: pOB7; Site 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGCAAGAG(G). Library constructed by Ling Hong in the  
 laboratory of Gerald M. Rubin (University of California,  
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
 Superscript II RT (Life Technologies). Note: this is a  
 NIH\_MGC Library."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 6.75e-133 Length: 922  
 Score: 1279.00 Matches: 275  
 Percent Similarity: 91.80% Conservative: 5  
 Best Local Similarity: 90.16% Mismatches: 16  
 Query Match: 60.19% Indels: 10  
 DB: 5 Gaps: 3

US-10-649-273-2 (1-414) x B0961028 (1-922)  
 Qy 80 ProProAlaIleGlnIleuHisArgGluAsnIleGlnArgIleValGlnGlnAlaLeu 99  
 Db 3 CTCACAGAGCTCAACAGCTTCAACAGAGAAATATTCAACGATAGTACAAAGCTCTT 62  
 Qy 100 SerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrlleYpsProGly 119  
 Db 63 TCTGCAAGTGGAGTCTCTCCAAAGTACCTCTCAGCAATTCACAACTCAATAAACCAGGA 122  
 Qy 120 LeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 139  
 Db 123 CTTCCTTTAAGCTCTGGAGTGGCTTATCATTTAGCTTACAGCTGTAGACAGTTAAAA 182  
 Qy 140 LysProPheIleProIleHisIleMetGluAlaHisAlaLeuThrlleArgLeuThraen 159  
 Db 183 AAGCATTCATTCATTCATATATATATGAGAGGCTCATGCACTTACTATTTAGTTGACCAAT 242  
 Qy 160 LysValGlnPheProPheLeuValLeuLeuIleSerGlyGlyHisCysleuLeuAlaLeu 179  
 Db 243 AAAGTGAATTTCTTTTACTTTTACTTTCTTTGATTTCTGAGAGTACAGTCTGTTGGCATTA 302  
 Qy 180 ValGlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAsp 199  
 Db 303 GTTCAAGAGGTTTCAAGATTTCTGCTTCTTGAAGTCTTTGGACATACACACAGGTGAC 362  
 Qy 200 MetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisIleProGlyCys-SerThrMet 219  
 Db 363 ATGCTTGAACAGGTGCAAGAGACCTTTCTTAATTAACATTCACAGTGCATTCACCAAT 422  
 Qy 219 tSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIle 239  
 Db 423 GAGTGTGGAAAGCCATAGAACATTTTGGCCAAACAGAAATAGATTTTCAATTTTACAT 482  
 Qy 239 eLysProProLeuHisIleAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisVal 259  
 Db 483 CAAACTCCCTTGCATCATGCTAAATATGTATTTTCTTTTACTGAGACTTCAACACCGT 542

QY 259 |ThrAsp|LyS|Ile|Met|Leu|Gly|Glu|Gly|Ile|Glu|Gly|Glu|Ile|Leu| 279  
 Db 543 |TACTGATTAATAATATATGAAAAAGAGAGAGATTTGAGAAAGGCAATCT 602  
 QY 279 |uSer|Ser|Ala|Ala|Ser|Ile|Ala|Ile|Val|Gln|His|Thr|Met|Ala|Cys|His|Ile|Val| 299  
 Db 603 |GTTCTTCAAGACGACATTCCTGCGACAGTACAGACACATGACATGCTGCTGAAA 662  
 QY 299 |sArg|Thr|His|Ser|Ala|Ile|Leu|Phe|Cys|Gln|Arg|Asp|Leu|Pro|Gln|Asn|Asn| 319  
 Db 663 |AAGAAA-CATCGGGCTATTCTCTTTGTAGACAGAGACTGTGACTCAAAATATATGC 721  
 QY 319 |aVal|Leu|Val|Ala|Ser-G|Gly|Val|Ala|Ser|Asn|Phe|Tyr|Ile|Arg|Arg|Ala|Leu|Glu| 339  
 Db 722 |AGTACTGGGTGATCTTTGGGGGGGTCCGACAGTAACTTCTATATCCCGCAAACTCTGAAA 781  
 QY 339 |Leu|Thr-Asn|Ala|Thr|Gln|Cys|Thr-Leu|Leu|Cys|Pro|Pro|Pro|Arg|Leu|Cys|Thr-A 358  
 Db 782 |AATTTCACAAACCCACACACGGGGCACCCTTTGTTGGGTCCCTCCCAACTATGCACTGG 841  
 QY 358 |sPAsn|Gly|Leu|Met|Ile|Ala|Trp-----Asn|Gly|Ile|Glu|Arg---Leu|Arg|Ala|Gly| 375  
 Db 842 |ATTAATGGG---CATTTATGATGGCTGGGGAGAGGAAATTAATAATCACTGCTCGGCT 898  
 QY 375 |eu|Gly 376  
 Db 899 |TGGGG 903  
 RESULT 7 701 bp mRNA linear EST 15-SEP-2000  
 BE740611 601595739F1 NIH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3949640 5',  
 LOCUS mRNA sequence.  
 DEFINITION BE740611.1 GI:10154603  
 ACCESSION BE740611.1 GI:10154603  
 VERSION EST.  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 701)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs-remail.nih.gov  
 Tissue Procurement: DCTD/DTF  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov  
 Plate: LNC814 row: n column: 09  
 High quality sequence stop: 701.  
 Location/Qualifiers  
 1..701  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3949640"  
 /issue\_type="adenocarcinoma cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_9"  
 /note="Organ: ovary; Vector: pOTB7; Site 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCACTGAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 FEATURES  
 source

ORIGIN  
 1  
 5

Alignment Scores:  
 Pred. No.: 3,84e-121 Length: 701  
 Score: 1173.00 Matches: 229  
 Percent Similarity: 99.15% Conservative: 3  
 Best Local Similarity: 97.86% Mismatches: 0  
 Query Match: 55.20% Indels: 2  
 DB: 2 Gaps: 0  
 US-10-649-273-2 (1-414) x BE740611 (1-701)

QY 125 |Gly|Val|Gly|Leu|Ser|Phe|Ser|Leu|Gln|Val|Gly|Gln|Leu|Gly|Pro|Phe|Ile|Pro 144  
 Db 2 |GGAATGGGCTTATATATTAGCTTACCTGCTGATGACAGTTAAACCAATTCATTTCTCC 61  
 QY 145 |Ile|His|Met|Glu|Ala|His|Ala|Leu|Thr|Ile|Arg|Leu|Thr|Asn|Val|Glu|Phe|Pro 164  
 Db 62 |ATTATCATATGAGGGCTCATGCACTTACTATTTAGTTGACCAATTAAGTAGAATTTCTCT 121  
 QY 165 |Phe|Leu|Val|Leu|Leu|Ile|Ser|Gly|His|Cys|Leu|Leu|Ala|Leu|Val|Gln|Val|Ser 184  
 Db 122 |TTTTAGTCTTTTGAATTTCTGAGGTCACTGCTGTTGGCATTAGTTCAAGAGATTCA 181  
 QY 185 |Asp|Phe|Leu|Leu|Leu|Gly|Val|Ser|Leu|Asp|Ile|Ala|Pro|Gly|Asp|Met|Leu|Asp|Val 204  
 Db 182 |GATTTCTGCTTCTTGGAAGCTTTTGACATACACAGGTGACATGCTTGACAAAGTGG 241  
 QY 205 |Ala|Arg|Arg|Leu|Ser|Leu|Ile|Val|His|Pro|Glu|Cys|Ser|Thr|Met|Ser|Gly|Val|Ala 224  
 Db 242 |GCAAGAAAGACTTCTTTTAATTAACATCCAGAGCTCCACACATGAGTGAGGAAAGCC 301  
 QY 225 |Ile|Glu|His|Leu|Ala|Val|Gln|Gly|Asn|Arg|Phe|His|Phe|Asp|Ile|Val|Pro|Leu|His 244  
 Db 302 |ATAGAACATTTGGCCCAACAAAGAAATGATTTGATTTGACATCAAACTCCCTTGCAAT 361  
 QY 245 |His|Ala|Val|Asn|Cys|Asp|Phe|Ser|Phe|Thr|Gly|Leu|Gln|His|Val|Thr|Asp|Val|Ile 264  
 Db 362 |CATGCTAAATAATGTGATTTTCTTTTAACTGACCTTCAACAGTTACGATTAATAATAATA 421  
 QY 265 |Met|Val---Lys|Glu|Val|Gln|Glu|Gly|Ile|Glu|Val|Gln|Ile|Leu|Ser|Ser|Ala|Ala 284  
 Db 422 |ATGAAACAGAGAAACAAAGAGAGATGAGAAAGGGGCAAAATCTGCTTCAGACAGCG 481  
 QY 284 |SPI|Lea|Ala|Leu|Val|Gln|His|Thr|Met|Ala|Cys|His|Leu|Val|Val|Val|Val|Val|Val 304  
 Db 482 |ACATTCCTGCGACAGTACAGACACATGCAATGCTGCTGTGAAAGAAACACATCGGG 541  
 QY 304 |Ile|Leu|Phe|Cys|Val|Gln|Arg|Asp|Leu|Pro|Gln|Asn|Asn|Ala|Val|Leu|Val|Ala|As 324  
 Db 542 |CTATTCTGTTTGTAGACAGAGACATTTGTAACCTCAAAATAATGACATGCTGTTGCAT 601  
 QY 324 |Arg|Gly|Val|Ala|Ser|Asn|Phe|Tyr|Ile|Arg|Val|Ala|Leu|Glu|Ile|Leu|Thr|Asn|Ala|T 344  
 Db 602 |CTGTGTGGTGTGCAAGTAATCTTATATCCGACAGGCTCTGGAATTTTAACAAACGCAA 661  
 QY 344 |Arg|Gln|Cys|Thr|Leu|Leu|Cys|Pro|Pro|Arg|Leu|Cys 356  
 Db 662 |CACAGTGACATTTGTTGTGCTCTCTCCACAGCATATGC 699  
 RESULT 8 749 bp mRNA linear EST 16-MAR-2004  
 CK982692 4115331 BARC 9BOV Bos taurus cDNA clone 9BOV45\_M02 5', mRNA  
 LOCUS sequence.  
 DEFINITION CK982692.1 GI:45500672  
 ACCESSION CK982692  
 VERSION CK982692.1 GI:45500672  
 KEYWORDS EST.  
 SOURCE Bos taurus (cow)  
 ORGANISM Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovinae; Bos.  
 1 (bases 1 to 749)  
 Sonstegard, T.S., Van Tassel, C.P., Matukumalli, L.K., Hartay,

TITLE  
G.P., Bosak, S., Rubenfield, M. and Gasbarre, L.C.  
Production of EST from cDNA libraries derived from immunologically  
activated bovine gut

JOURNAL  
Unpublished (2004)

COMMENT  
Contact: Ted S. Sonstegard  
Bovine Functional Genomics Laboratory  
Animal and Natural Resources Institute  
Bldg. 200 Rm2A BARC-East, Beltsville, MD 20705, USA  
Tel: 3015048416  
Fax: 3015048414  
Email: tadesant@barc.arsa.gov

Single pass sequencing. Bases called and trimmed with phred  
0.000925 using options -trim alt -trim fasta. Vector identified  
by cross match using options -mismatch 12 -mismatch 18  
Plate: 45 row: M column: 02  
Seq primer: CCCAGTCACGACCTGTGTAACG  
High quality sequence stop: 749.

## FEATURES

source

Location/Qualifiers  
1. 749  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/strain="Holstein"  
/db\_xref="taxon:9913"  
/clone="9BOV45\_M02"  
/sex="Male"  
/tissue\_type="Pooled"  
/dev\_stage="Multiple"  
/lab\_host="DH10B T1 phage resistant"  
/clone\_lib="BARC 9BOV"  
/note="Organ: Abomasum; Vector: pagen-1; Site 1: EcorV;  
Site 2: NotI; Equimolar amounts of mRNA extracted from  
fundic and pyloric abomasums of 18 and 21 week old steers.  
Exposure to Osteragia osteragi was initiated at 15 weeks  
of age. fundic and pyloric abomasum"

## ORIGIN

## Alignment Scores:

pred. No.: 2.09e-119 length: 749  
Score: 1158.00 Matches: 225  
Percent Similarity: 95.18% Conservative: 12  
Best Local Similarity: 90.36% Mismatches: 12  
Query Match: 54.49% Indels: 0  
DB: 7 Gaps: 0

US-10-649-273-2 (1-414) x CK982692 (1-749)

QY 49 ThrAlaAlaValValAspGluThrGlyAsnValLeuGlyAlaAlaLeHisSerGln 68  
DB 1 ACAGCAGCTCTGTAGTGAAGCGGAAATGTTTGGAGAGCAATCATTCCTCAA 60  
QY 69 ThrGluValHisLeuValsthrGlyTleValProProAlaAlaGlnGlnLeuHisArg 88  
DB 61 ACTGAAGTCATTAAACAGGCGGAGATTATCTCCAGAGCTCAACAGCTTCATAGA 120  
QY 89 GluAsnIleGlnAgtIleValGlnGlnAlaLeuSerAlaSerGlyValSerProSerasp 108  
DB 121 GAAATATTTCAACCCATAGTACAGAAAGCTCTCTCTGCAATGAGTCTTCCAAAGTAA 180  
QY 109 LeuSerAlaIleAlaThrThrIleValProGlyLeuAlaLeuSerLeuGlyValGlyLeu 128  
DB 181 CTCACAGCTGACACCACTATTAAGCCAGAGCTTCTTAAAGCTTGCGGTAGTTTA 240  
QY 129 SerPheSerLeuGlnLeuValGlyGlnLeuValysPhePheIleProIleHisMet 148  
DB 241 TCGTTTAGCTTACACACTGATGACCAAGTTTAAAAAGCCCTTCATTCATTCATCAGATG 300  
QY 149 GluAlaHisAlaLeuThrIleArgLeuThrAsnValGluPheProPheLeuValLeu 168  
DB 301 GAGGCTCATGACACTCATATAGGTTAACAATAGAGTAGAATTTCCGTTTATGTTCTT 360  
QY 169 LeuIleSerGlyGlyHisGlyLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 188  
DB 361 TTGATTCTGAGGTCATTGCTTTTGCACTAGTAGAGAGTTTCAGATTCTTCTT 420

QY 189 LeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspValAlaArgArgLeu 208  
DB 421 CTGGGAAGTCTTTGGACATAGCACAGGTACATGCTTGCAAGTAGTGAAGAAGACTT 480  
QY 209 SerLeuIleValHisProGluGlySerThrMetSerGlyGlyValAlaIleGlnHisLeu 228  
DB 481 TCCTTAATTAACATCCAGAGTGTCTCCACATGAGTGGCGGAAAGGCTATAGAACATTGG 540  
QY 229 AlAlaGlnGlyAsnArgPheHisPheAspIleLysProPheLeuHisAlaLysAsn 248  
DB 541 GCCAAACAGGAATATGATTCATTTCATTTCCAGCTCCCAAGCAACGCTTAAAT 600  
QY 249 CysAspPheSerPheThrGlyLeuGlnHisValThrAspValIleIleMetLysArgGlu 268  
DB 601 TGTATTTTCTTTTCTTGCACTCAACACGTTATGTAAGATGATATGACANNAAGGAA 660  
QY 269 LysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaThr 288  
DB 661 AAGAGGAAGGTATTCGACAGGCGGACAGGTCTGCTTCAGTCGACAGACATTGCTGCG 720  
QY 289 ValGlnHisThrMetAlaCysHisLeu 297  
DB 721 GTGACGACACCGTGGCTGCGCACTT 747

## RESULT 9

CO738006

LOCUS 800 bp mRNA linear EST 29-JUL-2004  
DEFINITION SLH03c19e02f1 squirrel heart library 1 Spermothilus lateralis cDNA  
clone 19e02 5', mRNA sequence.

ACCESSION

CO738006

VERSION

CO738006.1 GI:50825276

KEYWORDS

EST.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Andrew R. Cossins  
Laboratory for Environmental Gene Regulation  
University of Liverpool  
School of Biological Sciences, The Biosciences Building, Crown  
Street, Liverpool, United Kingdom, L69 7ZB  
Tel: +44(0)151-795-4510  
Fax: +44(0)151-795-4431  
Email: cossins@liv.ac.uk  
Vector has been trimmed from this EST.  
Plate: 19 row: e column: 02  
Seg primer: pf1c T7 (5'-AATACGACCTCATATAGG-3')  
High quality sequence stop: 800.

## FEATURES

source

Location/Qualifiers  
1. 800  
/organism="Spermothilus lateralis"  
/mol\_type="mRNA"  
/db\_xref="taxon:76772"  
/clone="19e02"  
/sex="Male & female"  
/tissue\_type="Heart"  
/dev\_stage="Adult"  
/lab\_host="E.coli B10ctromax DH10B"  
/clone\_lib="squirrel heart library 1"  
/note="Vector: pf1c; Site 1: SalI GTGACG; Site 2: BamHI  
GGATCC; Normalized and subtracted cDNA library prepared  
from heart of hibernating and summer animals"

## ORIGIN

Alignment Scores:

Pred. No.: 8,74e-118 Length: 800  
 Score: 1144.00 Matches: 226  
 Percent Similarity: 97.05% Conservative: 4  
 Best Local Similarity: 95.36% Mismatches: 7  
 Query Match: 53.84% Indels: 0  
 DB: 7 Gaps: 0

US-10-649-273-2 (1-414) x CO38006 (1-800)

QY 1 MetLeu1leuThrIlyThrAlaGlyValPhePheIysProSerIysArgIysValTyr 20  
 DB 88 ATGCTTAATTTGAAATTAAGACAGCGCGAGTTTCTTTTAAACATCCAAAGAAAGTTGAT 147  
 QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIlyValLeu 40  
 DB 148 GCATTTTAAAGAGTTTAAATTTTCACTCTGGAACACTATTTCTTCAATAAATGATATG 207  
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyValAsnVal 60  
 DB 208 GGATTTGAAACAGCTGTGATGATACAGCAGCTGCGGTGTGATGAACTGGAAATGTG 267  
 QY 61 LeuGlyValAlaIleHisSerGlnThrGluValHisIleuIysThrGlyIleValPro 80  
 DB 268 TTGGGGAAGCAATACATTTCCAAACTGAACTTCAATTAAACAGGTGGATTAATCTCT 327  
 QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
 DB 328 CCAGTAGCTCAACACTTCTCATAGAGAAACATTCACAGATATGACAGAGCTTTGT 387  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleIysProGlyLeu 120  
 DB 388 GCCAGTGGAAATATCTCCAGTAGTAACTCTCAGCATTCACATTAAGCCAGGACTT 447  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuIys 140  
 DB 448 GCTTTAAGCTTAGAGAGGAGCTTATCATTTAGTTTACGTGTGAGACCGAGTGAAG 507  
 QY 141 ProPheIleProIleHisIleHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnIys 160  
 DB 508 CCATTCATCCCATTCATCATCATATGAGAGCTCAGCGACTTCATTAAGTTGACCAATAA 567  
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
 DB 568 GTAGAGTTTCCATTTTATTTGATTTCTTGAGAGTCACTGCTCTTGAGCATTTAGT 627  
 QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyIysSerLeuAspIleAlaProGlyIysMet 200  
 DB 628 CAAGAGTTTCAGATTTTCTGCTTCTTGGAAGCTTTTGACATTCGACACAGGTGACATG 687  
 QY 201 LeuAspIysValAlaArgArgLeuSerLeuIleIysHisProGluCysSerThrMetSer 220  
 DB 688 CTTGACAGAGTACAGAGAGACATTTCTTTAATCAACATCCAGAAATCTCCACCATATG 747  
 QY 221 GlyIlyValAlaIleGlnHisIleuAlaIysGlnGlyAsnArgPheHisPhe 237  
 DB 748 GCGGGGAAGGCTATAGAACATTTGGCCAAACAGAAACAGATTTTCAATTT 798  
 RESULT 10  
 BM907988 1109 bp mRNA linear EST 12-MAR-2002  
 LOCUS AGENCOURT 6707465 NIH\_MGC\_119 Homo sapiens cDNA clone IMAGE:5744334  
 DEFINITION 5', mRNA sequence.  
 ACCESSION BM907988  
 VERSION BM907988.1 GI:19358367  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eureleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1109)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)

## COMMENT

Contact: Robert Strausberg, Ph.D.  
 Email: cgsbds-remail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLN at:  
 http://image.llnl.gov  
 plate: LLAM12767 row: b column: 07  
 High quality sequence stop: 602.

## FEATURES

## SOURCE

1. 1109  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5744934"  
 /issue\_type="medulla"  
 /lab\_host="DH10B"  
 /clone\_1lb="NIH\_MGC\_119"  
 /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: EcoRV (destroyed); RNA source normal medulla from  
 anonymous male age 27. Library is oligo-dT primed and  
 directionally cloned (EcoRV site is destroyed upon  
 cloning). Average insert size 1.3 kb, insert size range  
 0.9-3 kb. Library is normalized and enriched for  
 full-length clones and was constructed by C. Gruber  
 (Invitrogen). Research Genetics tracking code 013. Note:  
 this is a NIH\_MGC Library."

## ORIGIN

## Alignment Scores:

Pred. No.: 3.28e-116 Length: 1109  
 Score: 1132.00 Matches: 241  
 Percent Similarity: 88.89% Conservative: 15  
 Best Local Similarity: 83.68% Mismatches: 23  
 Query Match: 53.27% Indels: 9  
 DB: 5 Gaps: 4

US-10-649-273-2 (1-414) x BM907988 (1-1109)

QY 1 MetLeu1leuThrIlyThrAlaGlyValPhePheIysProSerIysArgIysValTyr 20  
 DB 129 ATGCTTAATTTGAAATTAAGACAGCGAGTTTCTTTTAAACATCCAAAGAAAGTTTAT 188  
 QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIlyValLeu 40  
 DB 189 GAATTTTAAAGAGTTTAAATTTTCACTCTGGAACACTATTTCTTCAATAAATGATATG 248  
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyValAsnVal 60  
 DB 249 GGAATTTAAACTAGTTTGTATGATGATACAGCACTCTGTGTGAGTGAAGAAATGTG 308  
 QY 61 LeuGlyValAlaIleHisSerGlnThrGluValHisIleuIysThrGlyIleValPro 80  
 DB 309 TTGGGAAGACAAATACATTTCCAAACTGAAGTTCAATTAAACAGGTGGATTTGCTCT 368  
 QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
 DB 369 CCAGACTCAACAGCTTCAACAGAAATATTCAGAAATGATACAGAAAGCTCTTTCT 428  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleIysProGlyLeu 120  
 DB 429 GCCAGTGAAGTCTCCAGAGTCACTCTGCAATATGCAATATCCATTAACACAGACTT 488  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuIys 140  
 DB 489 GCTTTAAGCTTAGAGAGGCTTATCATTTAGTTTACGTGTGAGACCGAGTGAAG 548  
 QY 141 ProPheIleProIleHisIleHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnIys 160  
 DB 549 CCATTCATTCATTCATCATATGAGAGCTCATCATTAATTAAGTTGACCAATTA 608



QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
DB 609 GFGAAATTCCTTTTATGTTCTTTGATTTCTGGAGGTCACTGTCGTGGCATTAAGT 668  
QY 181 GInGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
DB 669 CAGGAGTTTCATGATTTCTGCTTCCTGGAAGCTTTGACATTAACCCAGGGGACATG 728  
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyCysSerThMetSer 220  
DB 729 CTTGAACAGGGGCGACAGAAATTTCTTTATTAACATCCAGGGGCTCCACCATAGG 788  
QY 221 GlyGlyLysAlaIleGlyHisLeu---AlaLysGInGlyAsnArgPheHisAsp---Asp 238  
DB 789 GGGGGGAAAGCATAAACCATTTGGGCGACACAGAAATAGATTTCTTTGGGAT 848  
QY 239 IleLysProProLeuHisHis-AlaLysAsnCysAspPheSerPheThrgly--LeuGln 257  
DB 849 CCAAACTCCCTTCATATGTAATAAAGGGAATTTTCTTTATGGAATTTCAA 908  
QY 258 HisVal---ThrAspLysIleIleMetLysLysGlyLysGlyGlyIleGlu----- 274  
DB 909 CACCGTTAACTGTTAAATATATATGAAAGGAAAGGAAAGGGAATTTGGA 968  
QY 275 ---LysGlyGlnIleLeu 279  
DB 969 AAAAGGGCAAAATCTCG 986

RESULT 11  
LOCUS CFI14247 765 bp mRNA linear EST 23-JUL-2003  
DEFINITION Shultzomic07498 Rat lung airway and parenchyma cDNA libraries  
Rattus norvegicus CDNA clone NA4935 5', mRNA sequence.

ACCESSION CFI14247.1 GI:33173962  
VERSION CFI14247  
KEYWORDS EST.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 765)  
AUTHORS Shultz,M.A., Zhang,L., Gu,Y.-Z., Baker,G.L., Fannuchi,M.V.,  
Padua,A.M., Gurske,W.A., Morin,D., Penn,S.G., Jovanovich,S.B.,  
Plopper,C.G., and Buckpitt,A.R.  
Gene expression analysis in response to lung toxicants: I.  
Sequencing and microarray development  
Am. J. Respir. Cell Mol. Biol. 30 (3), 296-310 (2004)

## TITLE

JOURNAL CONTACT: Shultz MA  
COMMENT Dept. of Molecular Biosciences, School of Veterinary Medicine  
University of California, Davis  
1311 Haring Hall, One Shields Avenue, Davis, CA 95616, USA  
Tel: 530 752 0793  
Fax: 530 752 4698  
Email: mshultz@ucdavis.edu

FEATURES  
1. 765  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="NA4935"  
/sex="male"  
/tissue\_type="airway or parenchyma"  
/dev\_stage="adult"  
/clone\_id="Rat lung airway and parenchyma cDNA libraries"  
/note="Organ: lung; Vector: pGEM-1izf(-); Site 1: Eco RI;  
Site 2: Not I; mRNA was isolated from microdissected rat  
lung airways and parenchyma tissues."

Alignment Scores:  
Pred. No.: 4,26e-114 Length: 765  
Score: 1111.00 Matches: 216  
Percent Similarity: 93.78% Conservative: 10  
Best Local Similarity: 89.63% Mismatches: 15  
Query Match: 52.28% Indels: 0  
Gaps: 0

US-10-649-273-2 (1-414) x CFI14247 (1-765)

QY 24 ArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeuGlyIleGlu 43  
DB 39 AAGAAATTAATGTTCAACCTTAGAGCATTTTTCATCTAAACCTGTCCTGGAAATGAA 98  
QY 44 ThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnValLeuGlyGlu 63  
DB 99 ACCAGCTGTGATGACACGACGCTGTGTGTGTGATGAAACTGAAATGTGCTGGAGAA 158  
QY 64 AlaIleHisSerGlnThrGluValHisLeuLysThrGlyGlyIleValProProAlaAla 83  
DB 159 GCACGTCACCTCCAGACGGAAGTCATCTGAAACAGGTGGGATTTCTCTCACTCGT 218  
QY 84 GInGlnLeuHisArgGluAsnIleGlnArgIleValGInGluAlaLeuSerAlaSerGly 103  
DB 219 CAACAGCTTCACAGAGAAATATTCAGCGAATAGTAGAAGACCTTTCTGCCAGTGG 278  
QY 104 ValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeuAlaLeuSer 123  
DB 279 GTCTCCCAACGACCTTCAGCAATTTGCAATCACTCAACCAAGGAGCTGCTTAAGC 338  
QY 124 LeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysProPheIle 143  
DB 339 CTGGGGGTGGCTTATCTTATAGCTTACAGCTAGTAATCACTTTAAAGCATTTATC 398  
QY 144 ProIleHisHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPhe 163  
DB 399 CCAATCCATCAATGAGAGCTCAGCGCTGACTATTAAGCTGACCCACAAGATCGAATT 458  
QY 164 ProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnLysVal 183  
DB 459 CTTTTTATGTTCTTTATATTTCTGGAGGCCACTGCTGTGGCGTTAGTTCAAGTGT 518  
QY 184 SerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLys 203  
DB 519 TCAATTTTCTGCTCTCGGAAGTCCCTGACATAGGCGCAGGACATGCTTACAAAG 578  
QY 204 ValAlaArgArgLeuSerLeuIleLysHisProGlyCysSerThMetSerGlyGlyLys 223  
DB 579 GTGGCAAGAAAGCTTTCTTTATCAACATCCAGAAATGTTCTACAAATGAGTGGGAAA 638  
QY 224 AlaIleGlyHisLeuAlaLysGlnGlyAsnArgPheHisAspAspIleLysProProLeu 243  
DB 639 GCTATAGAAACATTTGGCCAAAGAGGAATAGATTTCACTTATCATTCACCCCATG 698  
QY 244 HisHisAlaLysAsnCysAspPheSerPheThrglyLeuGlnHisValThrAspLysIle 263  
DB 699 CAGATGTCTAAGACATGATTTTCTTTAAGGAGCTTCAACATGTCACGATAGCTA 758  
QY 264 Ile 264  
DB 759 ATA 761

RESULT 12  
LOCUS BO636028 640 bp mRNA linear EST 15-JUL-2002  
DEFINITION hd03d11.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he  
Homo sapiens CDNA clone hd03d11 5', mRNA sequence.  
ACCESSION BO636028  
VERSION BO636028.1 GI:21760487  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens



REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT  
FEATURES  
source

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 640)  
Wistow G., Bernstein S.L., Wyatt M.K., Ray S., Behal A., Touchman J.W., Bouffard G., Smith D. and Peterson K. Expressed sequence tag analysis of human retina for the NEURBank Project: Rebinding, an abundant, novel retinal cDNA and alternative splicing of other retina-preferred gene transcripts  
Mol. Vis. 8 (4), 196-204 (2002)  
22103461  
12107411  
Contact: Wistow G  
Section on Molecular Structure and Function  
National Eye Institute  
6/331, NIH, Bethesda, MD 20892-2740, USA  
Tel: 301 402 3452  
Fax: 301 496 0078  
Email: gdraeme@helix.nih.gov  
Plate: 03 row: d column: 11  
Seq primer: M13Rpl reverse primer (AB1).  
Location/Qualifiers  
1..640  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="hd03d11"  
/issue\_type="Retina"  
/dev\_stage="Adult"  
/lab\_host="EMDH10B"  
/clone\_11b="Human Retina cDNA (Un-normalized, unamplified): Hd/he"  
/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue was dissected from two 80 year old donors with no observed eye disease. 100ug of total RNA was used for library construction. A directionally cloned cDNA library in the pSPORT1 vector (Life Technologies) was constructed at Bioserve Biotechnology (Laurel MD) essentially following the protocols of the Superscript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor  
[5'-pGACTAGTCTTAGATCGCGCGCCGCC(7)15-3']. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:  
Pred. No.: 1..22e-112 Length: 640  
Score: 1097.00 Matches: 209  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 51.62% Indels: 0  
Gaps: 0

US-10-649-273-2 (1-414) x BQ636028 (1-640)

ORIGIN

206 ArgAGTLeuSerleuilelyshisproglucysSerThetsergllylysaiaile 225  
|||  
3 AAGAGCTTCTTCTTAATAAACAATCCAGAGTGTCCCACTGAGTGTGGAAAGCCATA 62  
|||  
226 GIuHtLeuAlaIysGlnIyAsnArGpPheHisPhaAspIlelySPProleuHtHis 245  
|||  
63 GAACATTTGGCCAAACAGGAATAGATTTCATTTGACATCAAACTCCCTGCATCAT 122  
|||  
246 AlAlYsAsnCYsaSPheSerPheThnGlyLeuGlnHisValThraSPlySileIleMet 265  
|||  
123 GCTAAAAATTGATTTCTTTTACTGCAGCTTCAACAGCTTCTGATAAATAATATG 182  
|||  
266 LysLysGlnLysGlnGlyIleGlyLysGlyGlnIleLeuSerSeraIaaIaSPile 285  
|||  
183 AAAAAAGAAAAAGAGAGATTTGAGAAAGGGCAAACTCTGCTTCAGACAGACAT 242  
|||

ORIGIN

Alignment Scores:  
Pred. No.: 1.08e-106 Length: 730  
Score: 1045.00 Matches: 204  
Percent Similarity: 93.51% Conservative: 12  
Best Local Similarity: 88.31% Mismatches: 15  
Query Match: 49.18% Indels: 0  
Gaps: 0

US-10-649-273-2 (1-414) x CN823245 (1-730)

ORIGIN

286 AlAlaThrValGlnHisThrMetAlaCYsHisLeuVallysaRgThrsArgAlaIle 305  
|||  
243 GCTGCCACAGTACAGCACAAGATGGCATTCCTTGGAAGAACACATCGGCTAAT 302  
|||  
306 LeuPheCYsLysGlnAysAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGly 325  
|||  
303 CTGTTTGTGAAGCAGAGAGACTGTACCTCAAAATATGAGTACGTTGCAATCGGT 362  
|||  
326 G1ValAlaSerAsnPheYrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGln 345  
|||  
363 GGTCTCCCAAGTAACTCTTATATCCGAGAGCTCTGGAATTTTAAACAAGCAACAG 422  
|||  
346 CysThrLeuLeuCYsProProProArgLeuCYsThraSPsngIyleMetIleAlaTrp 365  
|||  
423 TGCACTTTGTTGTCTCTCCCTCCAGACTATGACATGATATATGATATGATGATG 482  
|||  
366 AsnGlyIleGlyLysArgAlaGlyIleGlyIleLeuHisAspIleGlyIleArg 385  
|||  
483 AATGTATTTGAAGAACACTACGCTGCTGGCATTTTACATGACATGAGAGCATCCGC 542  
|||  
386 TyrGluProLysCYsProLeuGlyValAspIleSerLysGluValGlyLysAlaSerIle 405  
|||  
543 TATGAACCAAAATATCTCTTGGAGTACATATCAAAAGAGTTGAGAGACTTCCATA 602  
|||  
406 LysValProGlnLeuLysMetGluIle 414  
|||  
603 AAGTACCAACAATTTAAATGAGAGATA 629  
|||

RESULT 13  
CN823245  
LOCUS  
DEFINITION  
Oa splbn\_04N08 M13reverse Sheep spleen/brain pSPORT1 library Ovis  
arles cDNA clone Oa\_splbn\_04N08 5', mRNA sequence.  
CN823245  
VERSION  
CN823245.1 GI:47951314  
EST.  
KEYWORDS  
SOURCE  
ORGANISM  
Ovis aries (sheep)  
Ovis aries  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Caprinae; Ovis.  
1 (bases 1 to 730)  
Gossner A. and Hopkins J.  
Ovine spleen/brain cDNA library  
Unpublished (2004)  
JOURNAL  
COMMENT  
Contact: J Hopkins  
Veterinary Biomedical Sciences  
University of Edinburgh  
Summerhall Square, Edinburgh, EH9 1QH.  
Email: j.hopkins@ed.ac.uk  
Plate: 04 row: N column: 08  
Seq primer: M13reverse  
High quality sequence start: 6  
High quality sequence stop: 550.  
Location/Qualifiers  
1..730  
/organism="Ovis aries"  
/mol\_type="mRNA"  
/db\_xref="taxon:9940"  
/clone="Oa\_splbn\_04N08"  
/clone\_11b="Sheep spleen/brain pSPORT1 library"  
/note="Vector: pSPORT1"

[illegible]

Fax: 3015048744  
 Email: rbaumann@nrl.barc.usda.gov  
 Single pass sequencing. Bases called and trimmed with phred  
 0.000925 using options '-trim\_alt''-trim\_fasta. Vector identified  
 by cross match using options -mismatch 12 -mnscore 18  
 Plate: 20      Row: 0      Column: 20  
 Seq primer: CCTATTAGTCGACCTCTATGACAC  
 High quality sequence stop: 661.  
     Location/Qualifiers  
     1..661  
     1..661

```

/Note="Organ: Intestine; Vector: pCMVSPORT6.1; Site:1;
NotI; Site:2; EcoRI; Normalized cow cDNA intestinal mRNA
library in pCMVSPORT6.1, constructed from equimolar mRNA
pools derived from 5 sources, 4 lactating intestinal,
neonatal intestinal 4/5 lactating, proximal Duodenum,
jejunum, distal ileum, colon, 1/5 Neonatal, proximal
Duodenum, jejunum, distal ileum"
```

[illegible]

Db 479 GGGAGCTCTTTGGACATAGCACCAAGGTGACATGCTTGACAGAGTAGACAGACTTTCC 538  
Qy 210 LeuilelyshisProglucySserThrmSerclyGlylysaalleglwhisLeuvala 229  
Db 539 TTTATTAATCATCAAGAGTCTCCACCATAGAGGCGGGAAGGTATAGAACATTTGGCC 598  
Qy 230 LysGlnGlyAsnArgPhehisPheaspIlelySproProleuwhisshisalyAsnGly 249  
Db 599 AAACAGGGAATATGATTCATTTGATTCACGCTCCATGCACGCTAAATAATGT 658  
Qy 250 ASP 250  
Db 659 GAT 661  
RESULT 15  
BQ433135 879 bp mRNA linear EST 24-MAY-2002  
LOCUS AGENCOURT\_7760756 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:6016098  
DEFINITION 5', mRNA sequence.  
ACCESSION BQ433135  
VERSION BQ433135.1 GI:21172211  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE NIH-MGC http://mgs.nci.nih.gov/  
AUTHORS 1 (bases 1 to 879)  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LLM13213 row: d column: 19  
High quality sequence stop: 674.  
Location/Qualifiers  
1. 879  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6016098"  
/tissue\_type="embryonal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 92"  
/note="Organ: testis; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 2.5 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC library."  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.39e-103 Length: 879  
Score: 1018.50 Matches: 210  
Percent Similarity: 94.67% Conservative: 3  
Best Local Similarity: 93.33% Mismatches: 9  
Query Match: 47.93% Indels: 3  
DB: 5 Gaps: 1  
US-10-649-273-2 (1-414) x BQ433135 (1-879)  
Qy 1 MetLeuileleuthrlysthrAlaGlyValPhePheLySproSerlySArglyValTyr 20  
Db 201 ATGCTAATCTTGAAGTCAAGAGGTTTATTAACATCAAAAGAAAGTTTAT 260  
Qy 21 GluPheLeuArgSerPheAsnPhehisProGlyThrLeuPheLeuHislyValleu 40

Db 261 GAATTTTAAAGAACTTTAATTTTCATCTCGAAGACTATTTCTTCATTAATAGTATG 320  
Qy 41 GlyileGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60  
Db 321 GGATTTGAATCTAGTTGTGATGATAGACAGAGCTGTGGTGATGAACGGAATGTG 380  
Qy 61 LeuGlyGluAlailehisSergInthrGluValhisLeuLySthrGlylyleValPro 80  
Db 381 TTGGAGAACCAATACATCTCCCAAACTGAAGTTCAATTAATAAAGGCGGATGTTCT 440  
Qy 81 ProhlaalagIngluileuwhisargGluAsnIleGlnArgIleValGlnGlualeuSer 100  
Db 441 CCAGCACTCAACAGCTTCACAGAGAAATATTCAGAAATAGTACAGAAAGCTTTCT 500  
Qy 101 AlaserGlyValSerProSerAspLeuSerAlailealThrThrIlelySproGlyLeu 120  
Db 501 GCCAGTGAAGTCTTCCAGAGTACCTTCAGCAATTCACATACATTAACCGAGACTT 560  
Qy 121 AlaLeuSerleuGlyValGlyLeuSerPheSerleuGlnleuValGlyGlnleuLylys 140  
Db 561 GCTTAAGCTGGAGAGTGGCTTATCATTTAGCTTACAGCTGTAGGACAGCTAAAGA 620  
Qy 141 ProPheileProIlehisshisemetGluAlaHisAlaLeuThrIleArgLeuThrAsnLy 160  
Db 621 CCATTCATTCCTCCATTCATCATATGAGAGCTCATGCACTTACTATAGGTGACCAATTA 680  
Qy 161 ValGluPheProPheLeuValleuLeuileSerclyGlyhisCysLeuLeuAlaleuVal 180  
Db 681 GTAGAAATTTCTTTTATGTTCTTTGATTTCTGAGAGTCACTGTCTGTGGCATTTAGT 740  
Qy 181 GlnGlyValSerAspPheLeuLeuGlyLySserLeuAspIleAlaProGlyAspMet 200  
Db 741 CAAGGATTTTCAGATTTTCTGCTTCCTTGAAGCTTTGGACATAGCACCGTGACATG 800  
Qy 201 -LeuAspLySValAlaArgArgLeuSerleuile---LysHis-ProGluCysSerThrm 219  
Db 801 GCTTACAGAGGTGGGCAAGAAAGACTTCTTANATANAACATTCCAAGAGCTCCCCC 860  
Qy 219 etserGlyGly 222  
Db 861 ATGGANAGGT 871

Search completed: June 17, 2005, 01:29:52  
Job time : 3564.8 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: June 16, 2005, 17:01:19 ; Search time 2288.07 Seconds  
(without alignments)  
4441.797 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_148\_414  
Perfect score: 1385  
Sequence: 1 MEAHATIRLTKVKEPFLV.....DISKEVGASIKVPLQMEI 267

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues  
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+.p2n.model -DEV=xlp  
-Q=/cm2\_1/USPTO.spool\_p/US10649273/runat\_15062005\_11416\_6043/app\_query.fasta\_1.1429  
-DB=EST -OPMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10649273\_@CGN\_1.1\_8076\_@runat\_15062005\_11416\_6043 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST: \*  
1: gb\_est1: \*  
2: gb\_est2: \*  
3: gb\_hic: \*  
4: gb\_est3: \*  
5: gb\_est4: \*  
6: gb\_est5: \*  
7: gb\_est6: \*  
8: gb\_g881: \*  
9: gb\_g882: \*  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID       | Description |
|------------|-------|-------------|--------|----|----------|-------------|
| 1          | 1263  | 91.2        | 870    | 5  | BQ423651 | AGENCOURT   |
| 2          | 1224  | 88.4        | 2284   | 3  | AK045669 | Mus muscu   |
| 3          | 1216  | 87.8        | 1622   | 3  | AK011265 | Mus muscu   |
| 4          | 1097  | 79.2        | 640    | 5  | BQ636028 | hd03d11.y   |
| 5          | 1075  | 77.6        | 852    | 3  | BX391919 | BX391919    |
| 6          | 1071  | 77.3        | 1605   | 3  | BC030671 | Mus muscu   |
| 7          | 1053  | 76.0        | 701    | 2  | BR740611 | 601595739   |
| 8          | 944   | 68.2        | 922    | 5  | BQ961028 | AGENCOURT   |
| 9          | 925   | 66.8        | 658    | 7  | CF362328 | 829596 MA   |

|    |       |      |      |   |          |           |
|----|-------|------|------|---|----------|-----------|
| 10 | 916   | 66.1 | 822  | 7 | CF257246 | phs008_90 |
| 11 | 838   | 60.5 | 637  | 7 | CF941819 | 4065407 B |
| 12 | 830   | 59.9 | 597  | 6 | CB272391 | ma157912  |
| 13 | 819.5 | 59.2 | 792  | 5 | BU403563 | 604138456 |
| 14 | 816   | 58.9 | 730  | 7 | CB823245 | Ca. sp1dn |
| 15 | 789   | 57.0 | 545  | 1 | AV602901 | AV602901  |
| 16 | 786   | 56.8 | 866  | 5 | BU127463 | 603114407 |
| 17 | 784   | 56.6 | 490  | 6 | CB852881 | UI-CF-FNO |
| 18 | 749   | 54.1 | 736  | 7 | CK365185 | CK365185  |
| 19 | 749   | 54.1 | 1173 | 6 | CD508917 | CD508917  |
| 20 | 746   | 53.9 | 723  | 5 | BU261251 | 603502215 |
| 21 | 722   | 52.1 | 812  | 5 | BU264689 | 603784202 |
| 22 | 717   | 51.8 | 634  | 2 | AM601179 | RC1-BT025 |
| 23 | 709   | 51.2 | 484  | 1 | AJ670918 | AJ670918  |
| 24 | 708   | 51.1 | 736  | 6 | CA057753 | CA057753  |
| 25 | 707.5 | 51.1 | 701  | 5 | BU621780 | UI-H-FL1  |
| 26 | 704   | 50.8 | 749  | 7 | CK982692 | 4115331 B |
| 27 | 698   | 50.4 | 696  | 5 | BU302606 | 603739448 |
| 28 | 692   | 50.0 | 579  | 7 | CK819035 | CK819035  |
| 29 | 691.5 | 49.9 | 789  | 5 | BU242187 | 603781023 |
| 30 | 683   | 49.3 | 909  | 5 | BX756548 | BX756548  |
| 31 | 682   | 49.2 | 548  | 7 | CO880741 | CO880741  |
| 32 | 671   | 48.4 | 682  | 2 | BB043703 | BB043703  |
| 33 | 668   | 48.2 | 706  | 5 | BU202465 | 603949052 |
| 34 | 666   | 48.1 | 869  | 5 | BX754527 | BX754527  |
| 35 | 647   | 46.7 | 919  | 7 | CF407294 | CH3H047_E |
| 36 | 646   | 46.6 | 878  | 5 | BX776940 | BX776940  |
| 37 | 645   | 46.6 | 533  | 4 | BM126453 | BM126453  |
| 38 | 638   | 46.1 | 1082 | 5 | BX359023 | BX359023  |
| 39 | 625   | 45.1 | 506  | 2 | BF415802 | UI-R-CA1  |
| 40 | 618.5 | 44.7 | 879  | 5 | BU256052 | BU256052  |
| 41 | 614   | 44.3 | 861  | 5 | BU246158 | 603746167 |
| 42 | 611   | 44.1 | 1171 | 5 | BU261605 | 603501763 |
| 43 | 604   | 43.6 | 424  | 1 | AA273921 | vb99c03_r |
| 44 | 602   | 43.5 | 413  | 1 | AA589724 | v178c10_r |
| 45 | 600   | 43.3 | 863  | 5 | BU376295 | 603808890 |

## ALIGNMENTS

RESULT 1  
BQ423651  
LOCUS  
DEFINITION BQ423651 870 bp mRNA linear EST 23-MAY-2002  
AGENCOURT 7790948 NIH\_MGC\_72 Homo sapiens CDNA clone IMAGE:6065828  
5' mRNA sequence.  
ACCESSION BQ423651  
VERSION BQ423651.1 GI:21118966  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 870)  
NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
CONTACT: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC/DCTD/DTF  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
http://image.llnl.gov  
Plate: LNLN1342 row: 1 column: 21  
High quality sequence stop: 710.  
Location/Qualifiers  
1. 870  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6065828"

FEATURES  
source

```

/issue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NH MGCC 72"
/notes="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

```

## ORIGIN

## Alignment Scores:

```

Pred. No.:      1.9e-136      Length:      870
Score:          1263.00      Matches:      248
Percent Similarity: 98.42%      Conservative: 1
Best Local Similarity: 98.02%      Mismatches: 2
Query Match:     91.19%      Indels:      2
DB:              Gaps:      0

```

US-10-649-273-2\_COPY\_148\_414 (1-267) x BQ423651 (1-870)

```

QY      1 MetGlaAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
Db      112 ATGGAGGCTCATCTACTATTAGTTGACCAATTAAGTAGAATTTCTTTTATGTT 171
QY      21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db      172 CTTTGTATTCTCGAGGTCATCTGCTGTGGCATTTAGTCAAGAGTTTCAGATTTTCTG 231
QY      41 LeuLeuGlyLysSerLeuAspIleLeaProGlyAspMetLeuAspLysValAlaArg 60
Db      232 CTCTCTTGGAAGCTCTTGGACATAGCACCGAGTGACATGCTTGCACAGGTGGCAAGAGA 291
QY      61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGlnHis 80
Db      292 CTTTCTTTTAAATCAATCCAGAGTCTCCACATGAGTGGTGGGAAAGCCATAGAACAT 351
QY      81 LeuAlaLysGlnGlyAsnArgPheHisAspIleLysProProLeuHisHisAlaLys 100
Db      352 TTGGCCAAACAAAGAAATATGATTTTCATTTGACATCAACCTCCCTTGCATCATGCTAAA 411
QY      101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLys 120
Db      412 AATGTGATTTTCTTTTACTGAGCTTCAACACCTTCTGATTAATAATATGAAAAAG 471
QY      121 GluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaIleAspIleAla 140
Db      472 GAAAAAGAGAGATTTGAAAGGGGCAATCTGCTTCAGAGACATGCTGCC 531
QY      141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
Db      532 ACGATACAGACACATGCGATGTCATCTGTGAAAAACAACATCGGCTATTCTGTTT 591
QY      161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 180
Db      592 TGTAAACAGAGACATCTTACCTCAAAATATATGCACTAGCTGATCGGTGCTG 651
QY      181 AlaSerAspPheThrIleArgArgAlaLeuGlnIleLeuThrAlaAlaThrGlnCysThr 200
Db      652 GCAAGTAACCTCATATACCGACAGCTCTGAAATTTTAAACAAACGACACATGCACT 711
QY      201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220
Db      712 TTGGTGTCTCTCTCCCAACATATGACATGATATAGCATTTATGATTCATGAAATGCT 771
QY      221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGlu 240
Db      772 ATTTGAAAGACTACGTGCTGCTGGCATTTTATACATACATAGAGGATCCGCTATGAA 831
QY      241 -ProLysCysPro-LeuGlyValAspIleSerLys 251
Db      832 ACCAAATGTGCTCTTGGAGATGACATATCAAAA 866

```

RESULT 2  
AK045669

LOCUS AK045669 2284 bp mRNA linear HTC 03-APR-2004  
DEFINITION Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length enriched library, clone:B230219017 product:similar to PUTATIVE SIMULOGYCOPTERASE TYPE 2 [Homo sapiens], full insert sequence.  
ACCESSION AK045669  
VERSION AK045669.1 GI:26337528  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE 1 Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)  
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)  
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitanishi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakauechi, S., Ikegami, T., Kaishiki, K., Fujimake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matshiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)  
4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)  
5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)  
6 (bases 1 to 2284)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Horii, P., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Karoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Mizuta, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken

## COMMENT

Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

Please visit our web site for further details.  
URL: <http://genome.gsc.riken.jp/>  
URL: <http://fantom.gsc.riken.jp/>  
Location/Qualifiers

## FEATURES

source

1. .2284  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM:DB:B230219017"  
/db\_xref="FANTOM:DB:B230219017"  
/clone="B230219017"  
/sex="male"  
/tissue\_type="corpora quadrigemina"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
221. .1465  
/note="unnamed protein product; putative  
similar to PUTATIVE SIMLOGICOPROTEASE TYPE 2 (Homo  
sapiens) (SPTR|Q9H4B0, evidence: FASTA, 80%ID, 100%length,  
match=1242)"  
/protein\_id="BAC32450.1"  
/codon\_start=1  
/db\_xref="GI:26337529"  
/translation="MLMRRTAGALPKPKSKVYGFRRSRVHRTLSCHKLVGIEF  
SCDDTGAAYVDENVTGELHSGTQVHHTGGIVPPVAQQLHRENQRIYERTLSAC  
RTPSDLSAIAITTKPLALSLGVLSFSQLVNPKPPIPHHMAHALTRLNK  
VEPFLVLISGHCCLALVQVSDPLLSLDIAPGMDKVARLSLHKPECT  
MSGKALBQAKDGNRFTINPMQNAKNCDFSTGLQHTIDLVTHKEBEGIERG  
OLSSADIDIAAIVOHATFCHLAKRTHRIIFCOKONLSPANAVLYVSGVASLYLR  
KALEIVANAFOCTLCPPRLCTDNGIMANGIERLRAGLVADVEDIRYEKCL  
GVDISREVAEAIKVPRLKML"

polya\_signal

2262. .2267

polya\_site

/note="putative"  
2284  
/note="putative"

## ORIGIN

## Alignment Scores:

Pred. No.: 2,75e-131 Length: 2284  
Score: 1224.00 Matches: 233  
Percent Similarity: 92.51% Conservatve: 14  
Best Local Similarity: 87.27% Mismatches: 20  
Query Match: 88.38% Indels: 0  
DB: 3 Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x AK045669 (1-2284)

QY 1 MetGUAhahisalaLeuthrilearGleuthrAnlysValGIuphePropheluVal 20  
DB 662 ATGAGGCTCAGCAGCTAGCTATTAGGCTCACCAATAAGAGAAATTCCTTTTAACTT 721  
QY 21 LeuLeuileseerGlyGlyhisCyleuLeuLaleuValGinglyValSerAspPhelu 40  
DB 722 CTTTGTATTTCTGGCGGCTGCTGCTGTGGCATTAGCCAAAGGTTTCCGATTTCTCG 781  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArgArg 60  
DB 782 CTCCTTGGAGAGCTTTGGACATAGCACAGCCGACATGCTTGACAGGTGGCAAGAA 841  
QY 61 LeuSerLeuileYshisProGlyCysSerTherMetSerGlyGlyValAlaIleGluHis 80  
DB 842 CTTTCTTATTAACAACATCCAGATGTTCTACATAGGTGTGGCAAAAGCTATAGAAC 901  
QY 81 LeuAlaLysGlnGlyAsnArgPhehisPheAspIleYshProProLeuHisHisAlaLys 100  
DB 902 TTGGCAGAAAGCGGAAATAGATTCATTTTACTATTCATCCACCTATCAGATCTAAG 961  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThraAspIleIleMetLysLys 120  
DB 962 AATTGCAATTTTCTTTCAGCGGACTTCACATATTACTGATAGAGTAAATACACACAA 1021

QY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140  
DB 1022 GAAAGAGAAAGAGCATGAGAGAGGCAAAATCTGTCATAGCTGCAGACATGCTCTCT 1081  
QY 141 ThrValGlnIsthMeAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
DB 1082 GCGGTACGACATGACACAGCTGCGACCTTCCGAAAGAACACATCGCGCTATTCTGTT 1141  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuAlaIleSerGlyVal 180  
DB 1142 TGCAGACGAGAAATTTGCTCTCTCCAGCTAACCGACAGATTAGTTGTATCGAGGTCTT 1201  
QY 181 AlaSerAspPheYrIleArgArgAlaLeuGlnIleLeuThrAsnAlaIsthGlnCysThr 200  
DB 1202 GCAAGTAACTTGTCATTCATCGAAAGCATTTGGAATTTGCGAAATGCAACCCAGTCAGC 1261  
QY 201 LeuLeuCySPProProProArgLeuCySPThraAspGlyIleMetIleAlaIsthAsnLys 220  
DB 1262 TTGTTGTCTCCACCTCCAAAGCTGTGCATGACATGCAATGATTCATGATGATGA 1321  
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrglu 240  
DB 1322 ATGAAAGATTACGTGCTGCTGCGCTTGGCGCTTTTACATGATGTAAGACATCCGATATA 1381  
QY 241 ProLysCySPProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260  
DB 1382 CCAAAATCTCTCTTGAGATGACATATTCAGAGAGTTGAGAGAGCTGCCATAAAGTA 1441  
QY 261 ProGlnLeuLysMetGluIle 267  
DB 1442 CCGGATTAATAAATGGCAGCTT 1462

## RESULT 3

AK011265

LOCUS

DEFINITION

Mus musculus 10 days embryo whole body cDNA, RIKEN full-length enriched library, clone:2610001M19 product:similar to PUTATIVE SIMLOGICOPROTEASE TYPE 2 (Homo sapiens), full insert sequence.

ACCESSION

AK011265

VERSION

AK011265.1

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

REFERENCE

AUTHORS

Carninci, P. and Hayashizaki, Y.

TITLES

High-efficiency full-length cDNA cloning

JOURNALS

Meth. Enzymol. 303, 19-44 (1999)

PUBMED

99279253

MEDLINE

10349636

REFERENCE

AUTHORS

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL

MEDLINE

20530913

PUBMED

11076861

| REFERENCE | AUTHORS   | TITLE   | JOURNAL | REFERENCE | AUTHORS  | TITLE  | JOURNAL |
|-----------|---|---|---------|-----------|--|--|---------|
| 4         | The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium   | Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001) |         | 5         | The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.   | Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length CDNs Nature 420, 563-573 (2002) (Bases 1 to 1622) |         |
|           | Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukuishi, Y., Furuno, M., Hanagaki, T., Harai, A., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojiima, Y., Komou, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishii, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, K., Sano, H., Saeki, D., Shibata, K., Shibata, Y., Shinigawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toy, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y. | Direct Submission   |         |           | Submitted (10-JUL-2000) Yoshihide Hayashizaki. The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  | Please visit our web site ( <a href="http://genome.gsc.riken.jp/">http://genome.gsc.riken.jp/</a> ) for further details.                   |         |
|           | CNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGACAGATCCAGACCTCTTTTCTTTTTTTTAA 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGACAGATTTCGAGTAATTAATTAATCCCCCCCCC 3']. cDNA was cleaved with XhoI and SacI. Cloning sites, 5' end: XhoI; 3' end: SacI. Host: SOBR.   |   |         |           | location/Qualifiers<br>1..1622<br>/organism="Mus musculus"<br>/mol_type="mRNA"<br>/strain="C57BL/6J"<br>/db_xref="PANTOM DB:261000IM19"<br>/db_xref="taxon:10090"<br>/clone="261000IM19"<br>/tissue="whole body"<br>/clone_lib="RIKEN full-length enriched mouse cDNA library"<br>/dev_stage="10 days embryo"<br>207..1451<br>/note="unnamed protein product; putative similar to PUTATIVE SIALOGLYCOPROTEIN TYPE 2 (Homo sapiens) (SPTR Q9H4B0, evidence: FASTV, 80%ID, 100%length, match=1242)"<br>/codon_start=1<br>/protein_id="BAB27506.1"<br>/db_xref="GI:12847276"<br>/translation="MTMLRRKRAKIIPRPSSKYVGLPRRSYVHRTISCHLVIGIETSCDPTGAADVDETNVLGEALSTQTVLKTGIVLPVAQDLHEINQRIVEFTLSACRITPSDSIAAITTKTGPIALSIGVLSFSLQVNQPKPFIPIHMEAHATLTITNKKVFEPFLILSGCHLALVGVSDFPLNGSLDIADQMIDKVARSLIKIECEGTMSGGKALEOLAKGNRNPHFTNPKNQANCDPFEGTGHITDILITHKEEGEEKRKQLISSADIQAAYOHATACHAKTHRTAIIIFCKKNLLSPANNVLYVSGYSNLTYIRKALKDIVANATCTTLCPRLPCDTNGNMIAMNGIERLAGGVLDVEDINEFKPCPLGVDSIRSEVAEAIIVPRLKVAL" |  |         |

| polyA_signal   | 1605..1610  | /note="putative" |
|--|---|------------------|
| polyA_site   | 1622  | /note="putative" |
| ORIGIN   |   |                  |
| Alignment Scores:  |   |                  |
| Pred. No.:   | 1,446-130   | Length: 1622     |
| Score:   | 1216.00   | Matches: 232     |
| Percent Similarity:                                      | 92.13%  | Conservative: 14 |
| Best Local Similarity:                                   | 86.89%  | Mismatches: 21   |
| Query Match:   | 87.80%  | Indels: 0        |
| DB:  | 3   | Gaps: 0          |
| US-10-649-273-2_COPY_148_414 (1-267) x AK011265 (1-1622) |   |                  |
| QY   | 1 MetGluAaHiaIaIaLeuThrIleaRgLeuThraAaLysValaGluPhEProPhleuVal      | 20               |
| Db   | 648 ATGAGAGGCTCAGCAGCTACATATTAGGCTCACCAATAAGTAGAATTCTTTTAGTT        | 707              |
| QY   | 21 LeuLeuIleSeRgLyGlyNHIScYsLeuLeuAlaLeuValGInGlyValSerAspPheLeu    | 40               |
| Db   | 708 CTTTGTGATTTCTGGGGGTCACTGCCCTGTGGCATTAAGTCCAAAGGTGTTCCGATTTCCGT  | 767              |
| QY   | 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaAArgAs     | 60               |
| Db   | 768 CTCCTTGGAAAGTCTTTGAGATACACACAGCGACATGCTTGAACAAGGTGGCAAGAGA      | 827              |
| QY   | 61 LeuSerLeuIleLysHISProGluCySerThrMetSerGlyGlyLysAlaIleGluHIS      | 80               |
| Db   | 828 CTTTCTTTAAATCAAAATCCAGAAAGTTCTTAACATGAGTGGTGAAGAAAGCTATAGAACAG  | 887              |
| QY   | 81 LeuAlaLysGInGlyAsnArgPheHISPheAspIleLysProProLeuHISaIaLys        | 100              |
| Db   | 888 TTGGCCAAAGACGGAAATAGATTCCATTTTACTATCAATCCACTATGACGAATGCTAAG     | 947              |
| QY   | 101 AsnCyAspAspSerPheThrGlyLeuGInHISValThraAspLysIleIleMetLysLys    | 120              |
| Db   | 948 AATTGCCATTTTCTTTTACACGGAGCTTCAACATATTACTGATTAAGCTTAATTAACACACAG | 100              |
| QY   | 121 GluLysGInGluGlyIleGluLysGInIleLeuSerSerAlaAlaAspIleAlaIa        | 140              |
| Db   | 1008 GAAAGAGAGAGAGCATGATGAGAGAGCAAACTCTGTATCATGCTCAACATTCGTGCT      | 106              |
| QY   | 141 ThrValGInHISThrMetAlaCySHISLeuValLysArgThrHISArgAlaIleLeuPhe    | 160              |
| Db   | 1068 GCGGTACAGCATGACAACAGCGGTGCCACTTGGGAAAGAACATCGCCCTATTCTGTTT     | 112              |
| QY   | 161 CysLysGInaArgAspLeuLeuProGlnAsnAlaValLeuValAlaSerGlyGlyVal      | 180              |
| Db   | 1128 TGCAAGACGAAAAATTTCTCTCTCTCCAGCTAAACGAGTATGATGATGAGAGGT         | 118              |
| QY   | 181 AlaSerAsnPheTyrlleaRgArgAlaLeuGInIleLeuThraAsnAlaThrGInCyThr    | 200              |
| Db   | 1188 GCAAGTAACTGTATATCCGAAAGACATTTGGAATTGTGCGAAATGCAACGAGTGCAG      | 124              |
| QY   | 201 LeuLeuCySProProProArgLeuCySPHISAspGInLysIleMetIleAlaIleAspGly   | 220              |
| Db   | 1248 TTGTTGTGTCACCTCCCAAGACTGTGACACTGACATAGCATATGATTCATGAGTAATGA    | 130              |
| QY   | 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHISAspIleGluGlyIleArgTyrglu    | 240              |
| Db   | 1308 AATTGAAGATTAACGTGCTGGCTGGCTGGCTTTTACATGATGATGAGAACATCCCATATGAA | 136              |
| QY   | 241 ProLysCySProLeuGlyValaAspIleSerLysGluValaGlyGluAlaSerIleLysVal  | 260              |
| Db   | 1368 CCAAAATATCTCTTGGAGTACATATCCAGAGAAATGTGCAGAGCTGCATATAAGTA       | 142              |
| QY   | 261 ProGlnLeuLysMetGluIle 267                                       |                  |
| Db   | 1428 CCGCATTTAAATAATGCACTT 1448                                     |                  |

B0636028  
LOCUS B0636028 640 bp mRNA linear EST 15-JUN-2002  
DEFINITION hd03dl1.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he  
ACCESSION B0636028 Homo sapiens cDNA clone hd03dl1 5', mRNA sequence.  
VERSION B0636028.1 GI:21760487  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 640)  
AUTHORS Wistow,G., Bernstein,S.T., Wyatt,M.K., Ray,S., Behal,A.,  
Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.  
TITLE Expressed sequence tag analysis of human retina for the NEIBank  
Project: Retbindin, an abundant, novel retinal cDNA and alternative  
splicing of other retina-preferred gene transcripts  
JOURNAL Mol. Vis. 8 (4), 196-204 (2002)  
MEDLINE 22103461  
PUBMED 12107411  
COMMENT Contact: Wistow G  
Section on Molecular Structure and Function  
National Eye Institute  
6/331, NIH, Bethesda, MD 20892-2740, USA  
Tel: 301 402 3452  
Fax: 301 496 0078  
Email: graeme@helix.nih.gov  
Plate: 03 row: d column: 11  
Seq primer: M3KRP reverse primer (ABI).  
Location/Qualifiers  
1..640  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="hd03dl1"  
/issue\_type="Retina"  
/dev\_stage="Adult"  
/lab\_host="EMDH10B"  
/clone\_11b="Human Retina cDNA (Un-normalized,  
unamplified): hd/he"  
/note="Organ: Eye; Vector: pSPORT1; Neutral retina tissue  
was dissected from two 80 year old donors with no observed  
eye disease. 100ug of total RNA was used for library  
construction. A directionally cloned cDNA library in the  
pSPORT1 vector (Life Technologies) was constructed at  
Bioserve Biotechnology (Laurel MD) essentially following  
the protocols of the Superscript Plasmid System full  
details of which are contained in the manufacturer's  
instruction manual (<http://www.lifetech.com/>). First  
strand synthesis was carried out using a Not I  
primer-adaptor  
[5'-PACTAGATTCTAGATCGAGCGCGCCGCC(7)15-3']. EST analysis  
was performed on the unamplified library at the NIH  
Intramural Sequencing Center (NISC)."  
ORIGIN  
Alignment Scores:  
Pred. No.: 2,94e-117 Length: 640  
Score: 1097.00 Matches: 209  
Percent Similarity: 100.008 Conservative: 0  
Best Local Similarity: 100.008 Mismatches: 0  
Query Match: 79.21% Indels: 0  
DB: 5 Gaps: 0  
US-10-649-273-2\_COPY\_148\_414 (1-267) x B0636028 (1-640)  
QY 59 ArgArgLeuSerLeuIleLysHisProGluCysSerTherMetSerGlyGlyAlaIle 78  
DB 3 AGAAGACTTTCTTAATTAACATCCAGAGTCTCCACATGATGGTGGGAAAGCCATA 62  
QY 79 GluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHis 98  
DB 63 GAACATTTGGCCAAACAAAGAAATAGATTTCATTTGACATCAAAACCTCCCTGCATCAT 122

QY 99 AlAlaLysAncYsaapPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMet 118  
DB 123 GCTAAATATGTGATTTTCTTTACTGACCTTCAACAGCTTACTGATTAATATATATG 182  
QY 119 LysLeuGluLysGluGluGluGlyGlnLysGlnLysLeuSerSerAlaAlaAspIle 138  
DB 183 AAAAAAGAAAAAGGAAAGGATTTAGAAAGGGCAATCTGTCTTCAGACGACGACATT 242  
QY 139 AAlaAlaThrValGlnHissthMetAlaCysHisLeuValLysArgThrHisArgAlaIle 158  
DB 243 GCTCCACAGTACGACACACATGATGATCTTGTGAAAGAAACATCGGCTATT 302  
QY 159 LeuPheCysLysGlnArgAspLeuLeuProGlnIleAsnAlaValLeuValAlaSerGly 178  
DB 303 CTGTTTGTAAAGAGAGAGACTTGTACTCAAAATATAGCAGTACTGGTGCATCTGCT 362  
QY 179 GlyValAlaSerAsnPheTyrlleArgArgAlaLeuGluIleLeuThrAsnAlaThrGln 198  
DB 363 GGTGTCCCAAGTAACTTATATATCCGACAGCTCTGGAAATTTTAAACAAAGCACACAG 422  
QY 199 CysThrLeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrp 218  
DB 423 TGCACTTGTGTGTCTCTCCACAGACTATGACATGATTAATGGATTAATGATTCATGG 482  
QY 219 AsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArg 238  
DB 483 AATGATTTGAAAGACATACGTCGCTGCTGGCATTTTACATGACATGAAGACATCCGC 542  
QY 239 TyrGluProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIle 258  
DB 543 TATGACCAAAATATGCTCTTGGAGTAGACATATCAAAAGAGTTGGAAGCTTCCATA 602  
QY 259 LysValProGlnLeuLysMetGluIle 267  
DB 603 AAGTACCACAAATTAATAATGAGATA 629  
RESULT 5  
EX391919/c 852 bp mRNA linear EST 28-APR-2004  
LOCUS EX391919 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens  
DEFINITION B0391919 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens  
ACCESSION B0391919  
VERSION B0391919.2 GI:46846154  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 852)  
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT On May 13, 2003 this sequence version replaced gi:30611736.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: [seqref@genoscope.cns.fr](mailto:seqref@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
1240.r  
For more information about this cluster, see  
[http://www.genoscope.cns.fr/cdnas=CS0BA1035ZEB07\\_CS03317\\_1fc=1240.r](http://www.genoscope.cns.fr/cdnas=CS0BA1035ZEB07_CS03317_1fc=1240.r)  
FEATURES  
source  
1..852  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DK001YB02"



/cell\_type="HELA CELLS COT 25-NORMALIZED"  
/cell\_line="HELA"  
/clone\_id="Homo sapiens HELA CELLS COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-cligo(CT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

# ORIGIN

## Alignment Scores:

| Pred. No.:             | 1 65e-114 | Length:       | 852 |
|------------------------|-----------|---------------|-----|
| Score:                 | 1075.00   | Matches:      | 208 |
| Percent Similarity:    | 100.00%   | Conservative: | 0   |
| Best Local Similarity: | 100.00%   | Mismatches:   | 0   |
| Query Match:           | 77.62%    | Indels:       | 0   |
|                        | 5         | Gaps:         | 0   |

US-10-649-273-2\_COPY\_148\_414 (1-267) x BK391919 (1-852)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnValGluPheProPheLeuVal 20  
Db 626 ATGGAGGCTCATCTACTACTATTAGCTTGACCAATAAGATTTCTTTTATGTT 567  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
Db 566 CTTTGATTCTTGAGGCTCACTGCTGTTGGCATTTAGTCAAGAGTTTCAGATTTCTG 507  
QY 41 LeuLeuGlyLeuSerLeuAspIleAlaProGlyAspMetLeuAspValAlaArg 60  
Db 506 CTTCTTGGAAGCTTTTGACATAGCACAGGTGACATGCTTGACAAAGTGGCAAGAGA 447  
QY 61 LeuSerLeuIleHisHisProGluCysSerThrMetSerGlyGlyValIleGluHis 80  
Db 446 CTTTCTTTAATAAATCCAGAGTCTCCACCATGATGTTGGGAAAGCATGAAATCAT 387  
QY 81 LeuAlaLeuGlnGlyAsnArgPheHisPheAspIleLeuAspProLeuHisHisAlaVal 100  
Db 386 TTGGCCAAACAAAGAAATAGATTTCATTTCATCACTCAACCTCCCTTGACATCATGCTAAA 327  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspValIleIleMetLeuVal 120  
Db 326 AATTGATTTTCTTTTACTGACATTCACACCTTCTGTAATAAATAATGAAAAAG 267  
QY 121 GluLeuGlyGluGlyIleGlyLeuGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140  
Db 266 GAAAAAGAGAGAGATTTGAGAGGGGCAATCTGTTTCAGACAGACATTCCTGCC 207  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValIleValThrIleValAlaIleLeuPhe 160  
Db 206 AAGGTACAGACACAAATGATGATCTTGTGAAAGAACACATCGGGCATTTCTGTTT 147  
QY 161 CysGlyGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 180  
Db 146 TGTAAACAGAGACACTGTTACTCTCAAAATAATGACATGCTGATGCTGCTGCTC 87  
QY 181 AlaSerAspPheThrIleArgArgAlaLeuGluIleLeuHisAsnAlaThrGlnCysThr 200  
Db 86 GCAAGTAACCTTATATCCGAGAGCTCTGGAATTTTAAACAAACGACACAGTGA 27  
QY 201 LeuLeuCysProProProArgLeu 208  
Db 26 TTGTTGTGCTCTCTCCACAGCTA 3

RESULT 6  
BC030671 1609 bp mRNA linear HTC 19-NOV-2003  
LOCUS BC030671 Mus musculus O-sialoglycoprotein endopeptidase-like 1, mRNA (CDNA  
DEFINITION clone IMAGE:1226118), containing frame-shift errors.  
ACCESSION BC030671  
VERSION BC030671.1 GI:21040459  
KEYWORDS HTC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

## REFERENCE AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 1609)  
Klausner, R.D., Collins, F.S., Peingold, E.A., Grouse, L.H., Derge, J.G.,  
Altschul, S.F., Zeeberg, B., Wagner, L., Sherman, C.M., Schuler, G.D.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Scheetz, T.E., Brownstein, M.J., Udell, T.B., Tothiyuki, S.,  
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mulvaney, S.J., Bosak, S.A., McKean, P.J.,  
McKernan, R.D., Malek, U.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huyl, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E.,  
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
2 (bases 1 to 1609)  
Strausberg, R.  
Direct Submission  
Submitted (20-MAY-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: Marcello Bento Soares, Ph.D.  
cDNA Library Preparation: M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
contact: [amadansystemsbio.org](mailto:amadansystemsbio.org)  
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha  
Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting

## REMARK COMMENT

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>  
Series: IRAX Plate: 66 Row: e Column: 10  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 21312463  
This clone has the following problem: frame shifted.

## FEATURES

source  
1..1609  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1226118"  
/tissue\_type="Thymus gland, mouse"  
/clone\_id="Soares\_thymus\_2ndMT"  
/lab\_host="DH10B"  
/note="Vector: pT73-Pac"

## ORIGIN

Alignment Scores:  
Pred. No.: 1.21e-113 Length: 1609  
Score: 1071.00 Matches: 211  
Percent Similarity: 83.52% Conservative: 12  
Best Local Similarity: 79.03% Mismatches: 20  
Query Match: 77.33% Indels: 24  
DB: 3 Gaps: 1

US-10-649-273-2\_COPY\_148\_414 (1-267) x BC030671 (1-1609)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
Db 665 ATGAGAGCTACCGACGACTATTAGGCTCACCAATAGAGATTCTTTTGTGTT 724  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
Db 725 CTTTGTGATTCTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 784  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
Db 785 CTCCTTGGGAGGCTTTGGACATAGCACGCGACCATGCTTGACAAAGTGCAGAGAGA 844  
QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80  
Db 845 CTTTCTTATTCAACACATCCAGATGTTCTTACATGATGCTGTAAGAAAGCTATAGAACG 904  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
Db 905 TTGGCCAAAGACGGAATAGATTCTCATTTTACTCATCATCACTTACAGAAATGCTAAG 964  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
Db 965 AATTGCAATTTTCTTTCACGGGACTTCAACATATTACTGATAAGCTAATAACACACAG 1024  
QY 121 GluLysGlnGlyGlyIleGlyLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140  
Db 1025 GAAAGAAAGAGAGGCTTGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1084  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
Db 1085 GCGGTCAGAGCTGACAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1144  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAlaValLeuValAlaSerGlyGlyVal 180  
Db 1145 TGCAGACGAGAAATTTGCTCTCTCCAGCTAACCCAGATTAGTGTATGAGAGTCTT 1204  
QY 181 AlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200  
Db 1205 GCAGATTAAGCTTGTATCCGAAAGCATTTGCAAAATGCAAGACGAGTGACG 1264  
QY 201 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaTyrAsnGly 220  
Db 1265 TTGTTGTCCTCCTTGGAGATATCCAGAGAGTTCAGATGATGATGATGATGATGATGAT 1315  
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGlu 240  
Db 1315 ----- 1315  
QY 241 ProLysCysProLeuGlyValAspIleSerLysGlyValGlyGluAlaSerIleLysVal 260  
Db 1316 ---TGAGTCTCTTGGAGATATCCAGAGAGTTCAGATGATGATGATGATGATGATGAT 1372  
QY 261 ProGlnLysMetGluIle 267  
Db 1373 CCGCATTAATAAATGCACTT 1393  
RESULT 7  
BE740611 701 bp mRNA linear EST 15-SBP-2000  
LOCUS 601595739P1 NIH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3949640 5',  
DEFINITION mRNA sequence.  
ACCESSION BE740611  
VERSION BE740611.1 GI:10154603  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 701)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strusberg, Ph.D.  
Email: cga@bbs-remail.nih.gov  
Tissue Procurement: DCTD/DBP  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LLCM814 row: n column: 09  
High quality sequence stop: 701.  
Location/Qualifiers  
1..701  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3949640"  
/tissue\_type="adenocarcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH\_MGC\_9"  
/note="Organ: ovary; Vector: pOTB7, Site 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
ORIGIN  
Alignment Scores:  
Pred. No.: 4.6e-112 Length: 701  
Score: 1053.00 Matches: 206  
Percent Similarity: 99.05% Conservative: 3  
Best Local Similarity: 97.63% Mismatches: 0  
Query Match: 76.03% Indels: 2  
DB: 2 Gaps: 0  
US-10-649-273-2\_COPY\_148\_414 (1-267) x BE740611 (1-701)  
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
Db 71 ATGAGAGCTACCGACGACTATTAGGCTCACCAATAGAGATTCTTTTGTGTT 130  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
Db 131 CTTTGTGATTCTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 190  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
Db 191 CTTTCTTATTCAACACATCCAGATGTTCTTACATGATGCTGTAAGAAAGCTATAGAACG 250  
QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80  
Db 251 CTTTCTTATTCAACACATCCAGATGTTCTTACATGATGCTGTAAGAAAGCTATAGAACG 310  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
Db 311 TTGGCCAAAGACGGAATAGATTCTCATTTTACTCATCATCACTTACAGAAATGCTAAG 370  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
Db 371 AATTGCAATTTTCTTTCACGGGACTTCAACATATTACTGATAAGCTAATAACACACAG 430  
QY 120 GluLysGlnGlyGlyIleGlyLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140  
Db 431 GGAAGAAAGAGAGGCTTGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 490  
QY 140 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuP 160  
Db 491 CCACATGACGACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 550  
QY 160 HeCysLysGlnArgAspLeuLeuProGlnAsnAlaValLeuValAlaSerGlyGlyVal 180

Db 551 TTGTGACGACGAGACTGTGTTACTTCAAAATAATGACGATGCTGTCATCTGCTGCTG 610  
 Qy 180 a1a1aseaenpethy71leargalaleu9uileuthrasna1athrG1ncyst 200  
 Db 611 TCGCAGTACTTCTATATCGCAGACTCTCGAAATTTTAAACGCAACAGTGTCA 670  
 Qy 200 hrleuLeuCySPtProProaArgLeuCyS 209  
 Db 671 CTTGTGTGTCCTCTCCAGACTATGC 699  
 RESULT 8  
 LOCUS BQ961028 922 bp mRNA linear EST 21-AUG-2002  
 DEFINITION AGENCOURT 8863711 NIH\_MGC\_18 Homo sapiens cDNA clone IMAGE:6423902  
 5', mRNA sequence.  
 ACCESSION BQ961028  
 VERSION BQ961028.1 GI:22376506  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 BUKARYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 922)  
 NIH-MGC http://mgc.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-remail.nih.gov  
 Tissue Procurement: DCTD/BTP/Gazdar  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMU)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LMU at:  
 http://image.llnl.gov  
 Plate: LLCM2603 row: d column: 15  
 High quality sequence stop: 584.  
 Location/Qualifiers  
 1..922  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6423902"  
 /issue\_type="large cell carcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_18"  
 /note="Organ: lung; Vector: pOTB7; Site: 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGACACGAG(G). Library constructed by Ling Hong in the  
 Laboratory of Gerald M. Rubin (University of California,  
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
 Superscript II RT (Life Technologies). Note: this is a  
 NIH\_MGC Library."  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 3.65e-99 Length: 922  
 Score: 944.00 Matches: 207  
 Percent Similarity: 89.45% Conservative: 5  
 Best Local Similarity: 87.34% Mismatches: 16  
 Query Match: 68.16% Indels: 10  
 DB: 5 Gaps: 3  
 US-10-649-273-2\_COPY\_148\_414 (1-267) x BQ961028 (1-922)  
 Qy 1 MetG1uA1a1h1s1a1euthr1learg1euthrasn1ysva1g1upheprophel1eua1 20  
 Db 207 ATGAGAGCTATGACCTTACTATTAGGTGACATTAAGAGATTTCTTTTATGATT 266  
 Qy 21 LeuLeu1leSerg1g1y1h1s1CyS1euleu1aleu1a1g1nc1y1Val1SerASPhe1eu 40  
 Db 267 CTTTGAATTTCTGAGAGTCACTGTCCTTGGCATTAGTTCAAGAGTTTCAGATTTCGTG 326

Qy 41 LeuLeuG1ly1ysS1er1euaSP11e1a1eProG1y1aSPhe1leuaSP1y1sVal1a1arg1rg 60  
 Db 327 CTTCTTGGAAGGTCTTTGACATATGACACAGGTGACATGCTTGAACAAAGTGGCAAGAGA 386  
 Qy 61 LeuSer1eulle1ySh1eProG1uCyS-SerTher1e1Ser1e1g1y1s1a1a1e1e1u1h1 80  
 Db 387 CTTTCTTTAAATMAAACATCCAGAGTGCNTCCACATGATG1GTGTGGAAAGCCATAGAAC 446  
 Qy 80 sleuA1a1yS1e1nc1y1a1sna1rph1e1sPhe1aSP11e1ySPtPro1euh1s1h1a1a1y 100  
 Db 447 TTGGCCAAACAGCAAAATAGATTTCATTTTGACATCAAACTCTTCCTTCATCATCTAA 506  
 Qy 100 sasnCyAspSPheSerPheThrg1yLeuG1nh1sVal1Thrasp1y1e1le1e1y1s1y 120  
 Db 507 AAATTGATTTTCTTTTACTGACCTTCAACAGCTTCACTGATTAATAATATATGAAAA 566  
 Qy 120 eG1u1yS1e1u1g1u1g1y11eG1u1yS1e1yG1n11e1e1uSer1e1a1a1aSP11e1a1 140  
 Db 567 GGA1AAAGAGGAAAGATTTGAGAAAGGCAAAATCTCTTCAGCAGCAGCATTTGCTGC 626  
 Qy 140 aThrVal1G1nh1sThr1e1a1a1CySh1s1e1uVal1ySArThr1h1sArG1a1a1e1e1uPh 160  
 Db 627 CACAGTACAGCACAAATGCGATGTCATCTGTGAAAGAAA-CATCGGGCTATTCTGTT 685  
 Qy 160 eCy1yS1e1n1arG1aSPhe1e1uSPtProG1naSPnaSP1a1aVal1e1uVal1a1aSer-G1y1y 180  
 Db 686 TTGTAACGACGAGACATTTGTACTCAAAATATATGCACTATGCGTGCATCTGGGGGG 745  
 Qy 180 a1a1aseaenpethy71leargalaleu9uileuthrAsna1athrG1ncys 199  
 Db 746 TCGCAGTAACTTCTATATCCCGCAACTCGAATAATTACAAACCAACGCGGC 805  
 Qy 200 Thr-LeuLeuCySPtProProaArgLeuCySThr-AspAsnG1y1e1e1e1e1a1a1rP- 218  
 Db 806 ACCTTGTGTGGTCCCTCCCAACTATGACATGAGATTAATGG--CATTAATGATGTC 862  
 Qy 219 -----AsnG1y11e1G1u1a1r---LeuArG1a1a1y1e1uG1y 229  
 Db 863 TGGGAAAGGAAATTTAAATAATACNTGCTCGGCTTGGG 903  
 RESULT 9  
 LOCUS CF362328 658 bp mRNA linear EST 25-AUG-2003  
 DEFINITION CF362328 MRC 3P1G Sus scrofa cDNA 3', mRNA sequence.  
 ACCESSION CF362328  
 VERSION CF362328.1 GI:34161882  
 KEYWORDS EST.  
 SOURCE Sus scrofa (pig)  
 ORGANISM Sus scrofa  
 BUKARYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.  
 REFERENCE 1 (bases 1 to 658)  
 Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.T., Fox,J.,  
 Wise,T.A., Noneman,D.J., Wray,J.E. and Keefe,J.W.  
 A second set of porcine ESTs from a pooled-tissue normalized  
 library  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4350  
 Email: smtth@mail.marc.usda.gov  
 Single pass sequencing. Bases called with phred v0.020425.c and  
 trimmed with the aid of the trim\_alt option. Vector identified with  
 cross match v0.990329.  
 Plate: SRG8015 row: H column: 12  
 Seq primer: TAGAAGCAGCTCAGG.  
 FEATURES  
 source  
 1..658  
 /organism="Sus scrofa"  
 /mol\_type="mRNA"  
 Location/Qualifiers



| QY  | Db   | Accession     | Source | Reference Authors | Title | Journal | Comment |
|---|--|---------------|--------|-------------------|-------|---------|---------|
| 635   | TTTGTGTCCTCCTCCAAAGCTGTGCACCACTAATGTTGATTCATTCAGTAATGGC  | 694           |        |                   |       |         |         |
| 221   | TTTGTGTCCTCCTCCAAAGCTGTGCACCACTAATGTTGATTCATTCAGTAATGGC  | 694           |        |                   |       |         |         |
| 221   | TTTGTGTCCTCCTCCAAAGCTGTGCACCACTAATGTTGATTCATTCAGTAATGGC  | 694           |        |                   |       |         |         |
| 695   | ATTGAAAGGTTCGGCAGCAGATGTGGTATTTTATTCACGTACGATGATCGATCGATACGAA  | 754           |        |                   |       |         |         |
| 241   | ProlyscysProlenGlyValAspIleSerIysGluValAlaGlyGluAlaSerIleIysVal  | 260           |        |                   |       |         |         |
| 755   | CCAAAAGCTCCCTTGGAATGATATTTTCCAAAGATTGAA-GAGGATTCATCAAAAGTG   | 813           |        |                   |       |         |         |
| 261   | ProGlnLeu 263  |               |        |                   |       |         |         |
| 814   | CCAAAGCTA 822  |               |        |                   |       |         |         |
| RESULT 11   |  |               |        |                   |       |         |         |
| CR941819  | 637 bp   | mRNA          | linear | EST 15-MAR-2004   |       |         |         |
| LOCUS   | CR941819/c   |               |        |                   |       |         |         |
| DEFINITION  | 40655407 BARC 10BOV Bos taurus cDNA clone 10BOV12_F24 3', mRNA   |               |        |                   |       |         |         |
| ACCESSION   | CR941819   |               |        |                   |       |         |         |
| VERSION   | CR941819.1   | GI:45456199   |        |                   |       |         |         |
| KEYWORDS  | EST.   |               |        |                   |       |         |         |
| SOURCE  | Bos taurus (cow)   |               |        |                   |       |         |         |
| ORGANISM  | Bos taurus   |               |        |                   |       |         |         |
| REFERENCE   | Eukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.  |               |        |                   |       |         |         |
| AUTHORS   | 1 (bases 1 to 637)<br>Sonsiegard, T.S., Van Tassel, C.P., Matukumalli, L.K., Harhay, G.P., Boase, S., Rubenfield, M., and Gasparre, L.C.   |               |        |                   |       |         |         |
| TITLE   | Production of EST from cDNA libraries derived from immunologically activated bovine gut  |               |        |                   |       |         |         |
| JOURNAL   | Unpublished (2004)   |               |        |                   |       |         |         |
| COMMENT   | Contact: Tad S. Sonsiegard<br>Bovine Functional Genomics Laboratory<br>Animal and Natural Resources Institute<br>Bldg. 200 Rm2A BARC-East, Beltsville, MD 20705, USA<br>Tel: 3015048416<br>Fax: 3015048414<br>Email: tads@anri.barc.usda.gov |               |        |                   |       |         |         |
| FEATURES  | Source   |               |        |                   |       |         |         |
| Source  | 1..637   |               |        |                   |       |         |         |
| Location/Qualifiers   |  |               |        |                   |       |         |         |
| /organism="Bos taurus"  |  |               |        |                   |       |         |         |
| /mol_type="mRNA"  |  |               |        |                   |       |         |         |
| /strain="Holstein"  |  |               |        |                   |       |         |         |
| /db_xref="taxon:9913"   |  |               |        |                   |       |         |         |
| /clone="10BOV12_F24"  |  |               |        |                   |       |         |         |
| /sex="Male"   |  |               |        |                   |       |         |         |
| /tissue_type="Pooled"   |  |               |        |                   |       |         |         |
| /dev_stage="Multiple"   |  |               |        |                   |       |         |         |
| /lab_host="DH10B T1 phage resistant"  |  |               |        |                   |       |         |         |
| /clone_lib="BARC 10BOV"   |  |               |        |                   |       |         |         |
| /note="Organ: Small Intestine; Vector: pagen-1; Site: 1; Ecorev: Site 2; Ncrl: Equimolar amounts of mRNA extracted from proximal jejunums of 18 and 21 wk old steers, and distal ileums of 14 day old calves; proximal jejunum exposed to C. oncophora for 3 and 6 weeks, and distal ileum exposed to C. parvum for 7 days" |  |               |        |                   |       |         |         |
| ALIGNMENT SCORES:   |  |               |        |                   |       |         |         |
| Pred. No.:  | 5.09e-87   | Length:       | 637    |                   |       |         |         |
| Score:  | 838.00   | Matches:      | 153    |                   |       |         |         |
| Percent Similarity:   | 92.47%   | Conservative: | 19     |                   |       |         |         |
| Best Local Similarity:  | 82.26%   | Mismatches:   | 14     |                   |       |         |         |
| Query Match:  | 60.51%   | Indels:       | 0      |                   |       |         |         |

DB: 7 Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x CK941819 (1-637)

OY    82 AAlaysgInglYaSaMrqPhehiSPhesApriLeLySPProPoleuhShIsAlALySaSn 101  
Db     632 GCCAAACAGGAAAATAGAATTGCATTTTGAATTCAGCTCCCAACAACGGCTAAAAAT 573  
OY    102 CySaSpHeSeRPhethrGlYleuGlnHiVaIThrAspLysIlleMeTylySGlu 121  
Db     572 TGTGATTTTTCTTTTTCTGCCACTTCACAACCGTTTGTAATAAGATAATGCCAAAAGAA 513  
OY    122 LysGIuglUGlYileGuLYSGlGLHlleuSerSerIAaIaspiJaleAalathr 141  
Db     512 AAAGAGGAAGTATCAGACAGGGCAGGCTCTGCTTCAGCTGCAGACATTCCTGGC 453  
OY    142 ValGlnHisThmEtAlAcYSHisLeuValLYsaRgthrhisArGaIalleuPheCyS 161  
Db     452 GTGCAGACACACCGTAGCGCTGCACCATTCGCAAAAAGAACACATCGTGCTCTCTGTCG 393  
OY    162 LysGIuArGaSPLeuLeuProGlnASmsnAlaValLeuValAlaserGIeLYVaIALa 181  
Db     392 AAGCAGAGAGGCGTTCTTAACGTCAAGATCACACCACTAGCTGTGATCTGCAGAGCGTCGCA 333  
OY    182 SerSnphETyrILeaRGARgaLEuGLulleuThrAsnaIatrnGInCYSThrLeu 201  
Db     332 AGTAACCTAIVATATCCGAAAGCCCTGGAATTTGTACCAAATGCACAACGACGCACTTTG 273  
OY    202 LeuCysPROProproArgLeuCYSThrASPasnGlylleMeClalaIrPaSNglyILE 221  
Db     272 CTGTGCCCGCCCCCAGACTTTGCACAGTGCACACGGCGTTAGATTCAGTGAATGGTGT 213  
OY    222 GluArGLEuaRGaLaGlYleuGlylleuHisaspilleGUgLYlearTYrgLUpro 241  
Db     212 GAAGACACTACGTGCTGCGCTGGCATTTTACACAAACAGAGGCATCCGCTAGCAACA 153  
OY    242 LysCysPROleuNGlyVALaspIIAserLysGLuValGYguilaSerLIelYvaIPro 261  
Db     152 AAAAGTCTCTTGAGATGATVATCAAAAAGATGGAGAGCTGTATVAAAAGTGCCA 93  
OY    262 GluLeuLYsmetGluile 267  
Db     92 AGATTAAAAATGAAGATT 75

RESULT 12 CB272391 linear EST 24-FEB-2003  
LOCUS CB272391 mRNA 597 bp  
DEFINITION ma157612.v1 McCarrey Eddy spermatocytes Mus musculus cDNA clone IMAGE:6445750 5' similar to TR:Q9VMD6 Q9VMD6 CG14231 PROTEIN. ;,  
                rRNA sequence.  
ACCESSION CB272391  
VERSION CB272391 GI:28462714  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 597)  
McCarrey,J., Eddy,M., Marra,M., Hillier,L., Clifton,S., Pape,D.,  
Martin,J., Wylie,T., Dante,M., Bowers,Y., Theising,B., Gibbons,M.,  
Ritter,E., Tagareishvili,R., Ronko,I., Maguire,L., Kennedy,S.,  
Bennett,J., Waterston,R. and Wilson,R.

NIEHS Mouse Unpublished (2002)  
Contact: McCarrey/Buddy NIEHS Mouse  
NIHNS Mouse Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library constructed and donated by J. McCarrey, Ph.D. (Southwest  
Foundation for Biomedical Research, Dept. of Genetics) - ex-cislon

done by E.M. Eddy, Ph.D. (National Institutes of Health, National  
Institute of Environmental Health Sciences).

MGI:2069710

Seq primer: Primer name ambiguous

High quality sequence stop: 419.

# FEATURES

source

1. 597

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="CD-1"

/db\_xref="taxon:10090"

/clone="IMAGE:6445750"

/sex="male"

/tissue\_type="Spermatocytes, pooled from multiple mice"

/dev\_stage="60 day"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="McCarrey Eddy spermatoctyes"

/note="Organ: testis; Vector: pBluescript SK+

(Stratagene); Site 1: XhoI; Site 2: EcoRI; CDNA oligo

dr-primed [5'-(GA)10-ACTAGCTCGAGTGTGTTT-3'] and

directionally cloned using 5' linkers 5'-ATTGGGACGAG-3'

and 5'-CTCGTCCG-3'. Size selection of >400bp material

gives average insert size ranging from 1-2 kb. Library was

mass excised (from lambda-UniZAP-XR) and resulting

single-stranded phagemids were prepped and transformed

into DH10B. Library contains 98% recombinants.

References: J. Androl. 20:635-639 and Gene 25:263-269.

Library constructed and donated by J. McCarrey, Ph.D.

(Southwest Foundation for Biomedical Research, Dept. of

Genetics); excision done by E.M. Eddy, Ph.D. (National

Institutes of Health, National Institute of Environmental

Health Sciences). Original lambda-based library is

available through ATCC, catalog #63422."

# ORIGIN

## Alignment Scores:

Pred. No.: 3,986-86 Length: 597

Score: 830.00 Matches: 162

Percent Similarity: 91.80% Conservative: 6

Best Local Similarity: 88.52% Mismatches: 15

Query Match: 59.93% Indels: 0

DB: 6 Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x CB272391 (1-597)

```

QY      1 MetGUAAlahisAlaLeuThrlleArgLeuThraAlaValGluPheProPheLeuVal 20
         |||
DB      49 ATGAGGCTCAGCAGCTAGCTATTAGCTCACAATTAAGTAAGTAATTCCTTTTATT 108
         |||
QY      21 LeuLeuLlserGlyGlyHisGlyLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
         |||
DB      109 CTTTGTATTTCTGGCGGTCTGCTGCTGTGGCATTAGCCCAAGGTGTTCCGATTTTCTG 168
         |||
QY      41 LeuLeuGlyLysSerLeuAspLleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
         |||
DB      169 CTCCTTGGGAAGCTTTTGACATGACACAGGCAATGCTTGACAGGTGCGCAAGAA 228
         |||
QY      61 LeuSerLeuLlserHisProGlyLysSerThrmSerGlyGlyValAlaGlnHis 80
         |||
DB      229 CTTTCTTATCAAACTCCAGATGTTCTTCAATGATGCTGGAAGAAAGTATGAACAG 288
         |||
QY      81 LeuAlaLysGlnGlyAsnArgPheHisPheAspLllysProProLeuHisAlaLys 100
         |||
DB      289 TTGGCCAAAGACGGAATTAATTCATTCTTCAATTCACCTATGCAGAAATGCTAAG 348
         |||
QY      101 AsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspLysLleLleMetLysLys 120
         |||
DB      349 AATTGCAATTTTCTTTCACGGGACTTCACATATTAATGCTAATTAACACACACAG 408
         |||
QY      121 GlnLysGlnGlyLysLysGlnGlnLysSerSerAlaAlaAspLleAlaLys 140
         |||
DB      409 GAAAGAAAGAAAGCAATTGAGAAAGGCGCAATCTGTATCATGCTGCGACATCTCTCT 468

```

```

QY      141 ThrValGlnHisThrMetAlaCyHisLeuValLysArgThrHisArgAlaLleLeuPhe 160
         |||
DB      469 GCGGTACAGCAATGACACAGCTGCACCTTCGAAAGAAACACATCGGCTATTCTGTT 528
         |||
QY      161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
         |||
DB      529 TGCAGCAGAAAAATTTCTCTCTCCACGTAACGACGATTAAGTTGATCTGAGAGTCTT 588
         |||
QY      181 AlaSerAsn 183
         |||
DB      589 GCAAGTAAC 597

```

## RESULT 13

BU403563 792 bp mRNA linear EST 27-NOV-2002

LOCUS 60413845661 CSEQCHN59 Gallus gallus cDNA clone CHEST967014 5', mRNA

DEFINITION

Sequence.

ACCESSION BU403563.1 GI:25772619

VERSION BU403563.1

KEYWORDS EST.

SOURCE Gallus gallus (chicken)

ORGANISM Gallus gallus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

Phasianinae; Gallus.

1 (bases 1 to 792)

Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,

Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.

A Comprehensive Collection of Chicken cDNAs

Curr. Biol. 12 (22), 1965-1969 (2002)

PUBMED 12445392

COMMENT Contact: Simon Hubbard

Department of Biomolecular Sciences

University of Manchester Institute of Science and Technology

(UMIST)

PO Box 88, Manchester, M60 1QD, UK

Tel: 0161208930

Fax: 01612360409

Email: Simon.Hubbard@umist.ac.uk.

Location/Qualifiers

1. 792

/organism="Gallus gallus"

/mol\_type="mRNA"

/strain="White Leghorn, Hixex"

/db\_xref="taxon:9031"

/clone="CHEST967014"

/dev\_stage="36"

/lab\_host="DH10B"

/clone\_lib="CSEQCHN59"

/note="Organ: limbs; Vector: pBluescript II KS(+); Site\_1:

EcoRI; Site\_2: NotI; This normalized library was

constructed from 1 million independent clones. cDNA

synthesis was initiated using an oligo(dT) primer, using

methylated C in the first strand synthesis reaction.

Following this first strand reaction, double-stranded cDNA

was blunted, ligated to NotI adapters, digested with

EcoRI, size-selected, and cloned into the NotI and EcoRI

compatible sites of a custom modified MCS of the

pBluescript (KS+) vector. The library was normalized in 2

rounds using conditions adapted from Soares et al., PNAS

(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6

(1996): 791, except that a significantly longer

reannealing hybridization was used."

# ORIGIN

## Alignment Scores:

Pred. No.: 1,016-84 Length: 792

Score: 819.50 Matches: 164

Percent Similarity: 78.99% Conservative: 39

Best Local Similarity: 63.81% Mismatches: 53

Query Match: 59.17% Indels: 3

DB: 5 Gaps: 1

US-10-649-273-2\_COPY\_148\_414 (1-267) x BU403563 (1-792)

```
OY 5 AAlaLeuThr11leArgLeuThrAsnValAlGluPheProPheLeuVal1leuLeu1Ser 24
Db 2 GCACTTACCACTACGACTGACAGCAAGTGAATTTCCCTTCTTAGTCTTTACTCTCC 61
OY 25 G1yG1yH1sCysLeuLeuAlaLeuVal1G1nG1yVal1SerAspPheLeuLeuG1yLys 44
Db 62 GGAAGGTCATGGCAT-CTGGCAGTAGACACGAGAGTTTCAATTTCTCTTGGTGGACAG 120
OY 45 SerLeuAsp11leAlaProG1yAspMetLeuAspLysVal1AlaArgArgLeuSerLeu1le 64
Db 121 TCCATGATATATGACACAGAGTGCATGCTGATAGAGTAGCAAGAGGCTCTCTTAGT 180
OY 65 LysH1sProG1yCysSerThrMetSerG1yLysAla1leG1nH1sLeuAlaLysG1n 84
Db 181 AAGCAACCCGAGGACGACAGCATGGCCGGGAGAAAGCAATGAGCACCTGGCTCAAAAC 240
OY 85 G1yAsnArgPheH1sPheAsp11leLysProProLeuH1sH1sAlaLysAsnCyAspPhe 104
Db 241 GGAAGATGGCAACAGTACACTTTCAGACTTCCATGCAACAGTATCTTAATCTGATTTT 300
OY 105 SerPheThrG1yLeuG1nH1sVal1ThrAspLys11le1leuLysLysG1yLysG1n 124
Db 301 TCTTCTCCGAGCTTCAGAGCTTGTCAACAAAGCCATCTTCGAAAGAAAGAAAGAA 360
OY 125 G1y11leG1nLysG1yG1n1leLeuSerSerAlaAlaAsp11leAla1leuValG1nH1s 144
Db 361 GGATATTCAGAAAGGAGAAATCTGCTCGTTAAGACATCCCTGCTGCACAGCAC 420
OY 145 ThrMetAlaCysH1sLeuVal1LysArgThrH1sArgAla1leLeuPheCysLysG1nArg 164
Db 421 GATAGTGGCTGCTATATATATCCAGCGGACACACCGACATGCTCTTCTGACAAAGAAC 480
OY 165 AspLeuLeuProG1nAsnAsnAlaVal1leuVal1AlaSerG1yLysAla1leAspAsn 184
Db 481 AGCATATTTATTCAAAACCTGCACTGCTGTTATCAGAGAGTTCAGAGTATCAG 540
OY 185 Tyr11leArgArgAlaLeuG1n1leuThrAsnAla1leuG1nCysThrLeuLeuCysPro 204
Db 541 TATATCAGAAAGAGTACGACTCTGGCAAAATGCAAAACGTTTCTTCTCTCTCT 600
OY 205 ProProArgLeuCysThrAspAsnG1y1leMet11leAla1leuAsnG1y1leG1nArgLeu 224
Db 601 CTTCAAGGCTGTCACCGATATGATGTTATGATTCATGAGATGCAATGCAATGAG 660
OY 225 ArgAlaG1yLeuG1y1leLeuH1sAsp11leG1y1leArg1y1leArg1y1leArg1y1le 244
Db 661 CGGCGAGGATGTGTAT-TTATACAGTACGATGAGCATCCGCTACGAAACCAAAAGCTCC 719
OY 245 LeuG1yVal1Asp11leSerLysG1yVal1G1yAla1leSer11leLysVal1Pro 261
Db 720 CTTGGAATG---ATATTCCAAAGAGTGAAGAGATCCATCAAAAGTGCA 767
RESULT 14
CN823245 730 bp mRNA linear EST 02-JUN-2004
LOCUS CN823245
DEFINITION Oa脾脏_04N08_M13reverse Sheep spleen/brain p5port1 library Ovis
ACCESSION CN823245
VERSION CN823245.1 GI:47951314
KEYWORDS EST.
SOURCE Ovis aries (sheep)
ORGANISM Ovis aries
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Caprinae; Ovis (to 730)
REFERENCE 1 (bases 1 to 730)
AUTHORS Gossner,A. and Hopkins,J.
TITLE Ovine spleen/brain cDNA library
JOURNAL Unpublished (2004)
COMMENT Contact: J Hopkins
```

Veterinary Biomedical Sciences  
University of Edinburgh  
Summerhall Square, Edinburgh, EH9 1QH.  
Email: j.hopkins@ed.ac.uk  
Plate: 04 row: N column: 08  
Seq primer: M13reverse  
High quality sequence start: 6  
High quality sequence stop: 550.  
Location/Qualifiers  
1..730  
/organism="Ovis aries"  
/mol\_type="mRNA"  
/db\_xref="taxon:9940"  
/clone="Oa脾脏\_04N08"  
/clone\_lib="Sheep spleen/brain p5port1 library"  
/note="Vector: p5port1"

## ORIGIN

Alignment Scores:  
Pred. No.: 2.3e-84 Length: 730  
Score: 816.00 Matches: 157  
Percent Similarity: 93.30% Conservative: 10  
Best Local Similarity: 87.71% Mismatches: 12  
Query Match: 58.92% Indels: 0  
DB: Gaps: 7

US-10-649-273-2\_COPY\_148\_414 (1-267) x CN823245 (1-730)

```
OY 1 MeG1uAlaH1sAlaLeuThr11leArgLeuThrAsnLysValG1uPheProPheLeuVal 20
Db 193 ATGAGAGCTATGACATCTTACTATGAGTTAACAAAATGAGTAGAATTTCCATTTTATGTT 252
OY 21 LeuLeu11leSerG1yG1yH1sCysLeuLeuAlaLeuVal1G1nG1yVal1SerAspPhe 40
Db 253 CTTTGATTTCTGGAGGTCATGCTTTTGCACTTACCTGAGAGATTCAGATTTCTT 312
OY 41 LeuLeuG1yLysSerLeuAsp11leAlaProG1yAspMetLeuAspLysVal1AlaArgArg 60
Db 313 CTTTCTGAAAGTCTTTTGACATATGACACCGGACATGCTTGACAGTAGCAAGAAAGA 372
OY 61 LeuSerLeu11leLysH1sProG1yCysSerThrMetSerG1yLysAla1leG1nH1s 80
Db 373 CTTTGCTTAATTAACATCCAGAGTCTCCACATGAGTGGGAGAGGCTATGAAACAT 432
OY 81 LeuAlaLysG1nG1yAsnArgPheH1sPheAsp11leLysProProLeuH1sH1sAlaLys 100
Db 433 TTGGCCAAACAGGAGAAATGATGATTTGATTTCCAACTCCCATGCAACGTGCTAA 492
OY 101 AsnCyAspPheSerPheThrG1yLeuG1nH1sVal1ThrAspLys11le1leuLysLys 120
Db 493 AATTGTAATTTCTTTTCTGACATTCATATGATATGATATGATATGATATGATATG 552
OY 121 G1yLysG1nG1y1leG1yLysG1yG1n1leLeuSerSerAlaAlaAsp11leAlaAla 140
Db 553 GAAABACAGGAGGATATCGACAGGCGGACATCTGCTTCAGTCGACATGCTGCT 612
OY 141 ThrValG1nH1sThrMetAlaCysH1sLeuVal1LysArgThrH1sArgAla1leLeuPhe 160
Db 613 GCACTCAGACACACCGAGCTGCGACATGTCNAAGAAACACATCGGCTATTTCTGTT 672
OY 161 CysLysG1nArgAspLeuProG1nAsnAsnAlaVal1leuVal1AlaSerG1yG1y 179
Db 673 TGCAGAGAGAGGCTGTTATCATCAAGTACAGCACTGATGTTATCTGAGAGC 729
RESULT 15
AV602901 545 bp mRNA linear EST 27-NOV-2001
LOCUS AV602901/c
DEFINITION AV602901 Bos taurus kidney fetus Bos taurus cDNA clone B1K1013A07
ACCESSION AV602901
VERSION AV602901.1 GI:9725227
KEYWORDS EST.
SOURCE Bos taurus (cow)
```

## ORGANISM

Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovinae; Bos.

## REFERENCE

1 (bases 1 to 545)  
Takesuga, A., Hirotsune, S., Itoh, R., Jirohono, A., Suzuki, H., Aso, H.  
and Sugimoto, Y.  
Establishment of a high throughput EST sequencing system using  
poly(A) tail-removed cDNA libraries and determination of 36,000  
bovine ESTs

## TITLE

Nucleic Acids Res. 29 (22), E108 (2001)

## JOURNAL

## MEDLINE

## PUBMED

## COMMENT

Contact: Yoshikazu Sugimoto  
Animal Genetics Division  
Shirakawa Institute of Animal Genetics  
Odakura, Nishigo, Nishi-Shirakawa, Fukushima 961-8061, Japan  
Tel: 81-248-25-5641  
Fax: 81-248-25-5725  
Email: kazusugi@cocoa.ocn.ne.jp

Single pass sequencing.  
This clone was obtained from a polyA-deleted cDNA library.  
Location/Qualifiers

## FEATURES

source

1..545

/organism="Bos taurus"

/mol\_type="mRNA"

/db\_xref="taxon.9913"

/clone="EIKI013A07"

/tissue\_type="kidney"

/dev\_stage="fetus"

/lab\_host="DH10B"

/clone\_lib="Bos taurus kidney fetus"

/note="Vector: pZL1; Site\_1: SalI; Site\_2: NotI; Poly A

was deleted from a NotI site"

## ORIGIN

## Alignment Scores:

| Pred. No.:             | 2,14e-81 | Length:       | 545 |
|------------------------|----------|---------------|-----|
| Score:                 | 789.00   | Matches:      | 147 |
| Percent Similarity:    | 90.11%   | Conservative: | 17  |
| Best Local Similarity: | 80.77%   | Mismatches:   | 17  |
| Query Match:           | 56.97%   | Indels:       | 1   |
| DB:                    | 1        | Gaps:         | 0   |

US-10-649-273-2\_COPY\_148\_414 (1-267) x AV602901 (1-545)

```

QY      85  G|AAsnArpPhehIsPheAspIleLyPProPLeuhIsHs-AlAlysaEnCyAspPh 104
DB      544  GGAATATGATTCATTTTGAATTCACANCTCCATGCACGCGTCTAAATTTGATTT 485
QY      104 eSerPheThGlyLeuGlnHsValThrAspLysIleIleMetLysLysGluGluG 124
DB      484  TTCCTTTTCGACTTCACACGTTATGATTAAGATGATTAATGCAAAAGAAANAGCA 425
QY      124 uGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleValGlnH 144
DB      424  AGGATATCGACAGGCGGCGCTGCTTCAGTCGCGACATTCGCTGTTCGCTCCAGCA 365
QY      144 sThrMetAlaCyHsIleuValLysArgThHsArgAlaIleLeuPheCyLysGlnAr 164
DB      364  CACCGTGGCTGCGACATGCAAAAGAACACATGCTCTTCGTTCTGCAAGCAGAG 305
QY      164 gAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPh 184
DB      304  AGGCTTCTTACATCAAGATTAAGCAGTACTGTTATCTGAGGCGCTCGCAATTAATT 245
QY      184 eTyrlLeArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCyThrLeuLeuCyPr 204
DB      244  ATATATCCGAAAGCCCTCGAAATTTGACCAATGCAACACAGTGCACCTTGTGTGCC 185
QY      204 oProProArgLeuCyThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLe 224
DB      184  GCCCCCACTGCACTGACATGCAACGCGGTATGATTCATGCAATGGAATGTTGAAAGACT 125

```

```

QY      224 uArgAlaGlyLeuGlyIleLeuHsAspIleGluGlyIleArgTyGluProLysCyPr 244
DB      124  ACGTGTGGCTTGGGCAATTTTACACACAGAAAGGCATCCGCTACGAACCAAAATGTC 65
QY      244 oLeuGlyValAspLysSerLysGlyValGlyGluAlaSerIleLysValProGlnLeu 264
DB      64  TCTTGAGTATGATATATCAAAAGAAAGTTGAGAAAGCTGCTATAAAGTCCAAAGTTAA 5
QY      264 sMet 265
DB      4  AATG 1

```

Search completed: June 17, 2005, 01:30:00  
Job time : 2296.07 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM protein - nucleic search, using frame\_plus.p2n model

Run on: June 16, 2005, 17:01:19 ; Search time 2048.13 Seconds  
(without alignments)  
4441.797 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414  
Perfect score: 1240  
Sequence: 1 L1ALVQGVSDPFLGKSLDI.....DISKEVGEASIKVPQUMKI 239

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 34239544 seqs, 19032134700 residues  
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_plus.model -DEV=xlp  
-O=/cgn2\_1/USPTO.spool.p/US10649273/runat\_15062005\_111416\_6043/app.query.fasta\_1.1429  
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=bl0sum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10649273@cgn2\_1.1.8076@runat\_15062005\_111416\_6043 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEX=7

## Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gsa1:\*  
9: gb\_gsa2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID       | Description         |
|------------|-------|-------------|--------|----|----------|---------------------|
| 1          | 1118  | 90.2        | 870    | 5  | BQ423651 | BQ423651 AGENCOURT  |
| 2          | 1057  | 88.5        | 840    | 5  | BQ636028 | BQ636028 hcd03d11.y |
| 3          | 1079  | 87.0        | 2284   | 3  | AK045669 | AK045669 Mus muscu  |
| 4          | 1071  | 86.4        | 1622   | 3  | AK011265 | AK011265 Mus muscu  |
| 5          | 930   | 75.0        | 852    | 5  | BX391919 | BX391919 Mus muscu  |
| 6          | 526   | 74.7        | 1609   | 3  | BCU30671 | BCU30671 Mus muscu  |
| 7          | 908   | 73.2        | 701    | 2  | BE740611 | BE740611 Mus muscu  |
| 8          | 876   | 70.6        | 658    | 2  | CF362328 | CF362328 829596 MA  |
| 9          | 838   | 67.6        | 637    | 7  | CK941819 | CK941819 4065407 B  |

|    |       |      |      |   |          |                    |
|----|-------|------|------|---|----------|--------------------|
| 10 | 799   | 64.4 | 922  | 5 | BQ61028  | BQ61028 AGENCOURT  |
| 11 | 789   | 63.6 | 545  | 1 | AV602901 | AV602901           |
| 12 | 784   | 63.2 | 490  | 6 | CB852881 | CB852881 UI-CF-FN0 |
| 13 | 783   | 63.1 | 822  | 7 | CF257246 | CF257246 phao08_g0 |
| 14 | 753   | 60.7 | 866  | 5 | BU127463 | BU127463 603114407 |
| 15 | 749   | 60.4 | 736  | 7 | CK365185 | CK365185 AGENCOURT |
| 16 | 746   | 60.2 | 723  | 5 | BU261251 | BU261251           |
| 17 | 731.5 | 59.0 | 792  | 5 | BU403563 | BU403563 604138456 |
| 18 | 717   | 57.8 | 634  | 2 | AM601179 | AM601179 RCL-BT025 |
| 19 | 708   | 57.1 | 736  | 6 | CA057753 | CA057753 ssa1rpb54 |
| 20 | 707.5 | 57.1 | 701  | 5 | BU621780 | BU621780 UI-H-FLL- |
| 21 | 698   | 56.3 | 696  | 5 | BU302606 | BU302606 603739448 |
| 22 | 685   | 55.2 | 597  | 6 | CB272391 | CB272391 ma157912  |
| 23 | 683   | 55.1 | 909  | 5 | BX756548 | BX756548 BX756548  |
| 24 | 682   | 55.0 | 548  | 7 | CO880741 | CO880741 Boygen_09 |
| 25 | 671   | 54.1 | 682  | 2 | BB043703 | BB043703 BB043703  |
| 26 | 671   | 54.1 | 730  | 7 | CN823245 | CN823245 Oa.sp1bn  |
| 27 | 668   | 53.9 | 706  | 5 | BU202465 | BU202465 603949052 |
| 28 | 666   | 53.7 | 869  | 5 | BX754527 | BX754527 BX754527  |
| 29 | 647   | 52.2 | 919  | 7 | CF407294 | CF407294 CH3#047_E |
| 30 | 646   | 52.1 | 484  | 1 | AJ670918 | AJ670918 AJ670918  |
| 31 | 646   | 52.1 | 878  | 5 | CD508917 | CD508917 CDA93-E05 |
| 32 | 642   | 51.8 | 1173 | 6 | CD508917 | CD508917 CDA93-E05 |
| 33 | 638   | 51.5 | 1082 | 5 | BX359023 | BX359023 BX359023  |
| 34 | 625   | 50.4 | 506  | 2 | BF415802 | BF415802 UI-R-CAI- |
| 35 | 614   | 49.5 | 861  | 5 | BU246158 | BU246158 603779906 |
| 36 | 611   | 49.3 | 1171 | 5 | BU261605 | BU261605 603501763 |
| 37 | 604   | 48.7 | 424  | 1 | AA273921 | AA273921 VB99C03_r |
| 38 | 600   | 48.4 | 863  | 5 | BU376295 | BU376295 603808990 |
| 39 | 594   | 47.9 | 389  | 4 | BM744822 | BM744822 K-EST0018 |
| 40 | 591   | 47.7 | 812  | 5 | BU246489 | BU246489 603784202 |
| 41 | 589   | 47.5 | 357  | 5 | BQ672554 | BQ672554 AGENCOURT |
| 42 | 581   | 46.9 | 704  | 5 | BU327284 | BU327284 603491570 |
| 43 | 576   | 46.5 | 826  | 7 | CR444994 | CR444994 CR444994  |
| 44 | 573   | 46.2 | 385  | 1 | AJ647827 | AJ647827 AJ647827  |
| 45 | 572.5 | 46.2 | 613  | 1 | AA920105 | AA920105 v225h05.r |

## ALIGNMENTS

RESULT 1  
BQ423651  
LOCUS  
DEFINITION  
AGENCOURT 7790948 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:6065828  
5', mRNA sequence.  
ACCESSION  
BQ423651  
VERSION  
BQ423651.1 GI:2118966  
KEYWORDS  
SOURCE  
EST.  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 870)  
NIH-MGC http://imgc.nci.nih.gov/.  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-rcmail.nih.gov  
Tissue Procurement: ATCC/CDT/DTF  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: L1AM13342 row: 1 column: 21  
High quality sequence stop: 710.  
Location/Qualifiers  
1. 870  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6065828"

## FEATURES

source

/issue\_type="melanotic melanoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGCC 72"  
 /note="Organ: skin; Vector: PCMV-SPORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally. primer: Oligo dT.  
 Average insert size 2 kb. Library constructed by Life  
 Technologies."

# ORIGIN

## Alignment Scores:

Pred. No.: 36-119 Length: 870  
 Score: 1118.00 Matches: 220  
 Percent Similarity: 98.22% Conservative: 1  
 Best Local Similarity: 97.78% Mismatches: 2  
 Query Match: 90.16% Indels: 2  
 DB: 5 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x BQ423651 (1-870)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlnGlySerLeuAspIle 20  
 DB 196 CTTTGGCATTACTTCAAGAGATTTCAGATTTCCTGTTTGGAAAGCTTTGGACATA 255  
 QY 21 AlaProGlyAspMetLeuAspIleValAlaArgArgLeuSerLeuIleValSHISProGlu 40  
 DB 256 GCACCAAGGACATGCTTGACAGGTGGCAAGAACTTTCTTAATTAATAACATCCAGAG 315  
 QY 41 CysSerThrMetSerGlyGlyValAlaIleGlnIleLeuAlaValGlnGlyValAsnArgPhe 60  
 DB 316 TGCTTCACCATGAGTGTGGTGGAAAGCCATTAAGAACTTTGGCCAAACAAAGAAATAGATT 375  
 QY 61 HisPheAspIleLeuSerProPheLeuHisAlaValAsnGlyValAspPheSerPheThrGly 80  
 DB 376 CATTGGACATCAAACTCCCTTGATCATGCTTAATAATTTGATTTTCTTTACTGGA 435  
 QY 81 LeuGlnHisValThrAspIleIleMetLeuValGlnGlyValGlnGlyValIleGluLys 100  
 DB 436 CTTCAACACGTTACTGATAAATATATATGAAAGAAAGAAAGAGATTTGAGAGAG 495  
 QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleValGlnHisThrMetAlaCys 120  
 DB 496 GGGCAAAATCTGCTTTCAGCAGCAGACATTTGCTCCACATACAGCACACAATGGCATGT 555  
 QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140  
 DB 556 CATCTTGTAAGAAACAACATCGGCTATTCTGTTTGTAGAGAGAGACCTGTACTCT 615  
 QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaIleSerAsnPheValIleArgArg 160  
 DB 616 CAAAATATGCACTACTGTTGATGCTGCTGCTGCCAAGTAATTTCTATATCCGAGA 675  
 QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
 DB 676 GCTCTGGAATTTTAAACAAACGACACAGTGCATTTGCTGCTCTCCACAGACTA 735  
 QY 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGlnArgLeuAlaGlyLeu 200  
 DB 736 TGCACATGATTAATGCACTTATGATTCATGCAATGTAATTAAGACATCACTGCTGCTG 795  
 QY 201 GlyIleLeuHisAspIleLeuGlyIleArgArgIleProLeuCysAspProLeuGlyValAla 220  
 DB 796 GGCATTTTACATACATTAAGAGCAATCCGCTATGAACCAAAATGCTCTCTTGGAGTAG 855  
 QY 220 spIleSerLys 223  
 DB 856 ACATATCAAAA 866  
 RESULT 2 BQ636028 640 bp mRNA linear EST 15-JUN-2002  
 BQ636028 hd03d11.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he  
 DEFINITION Homo sapiens cDNA clone hd03d11 5', mRNA sequence.  
 ACCESSION BQ636028

# VERSION

BQ636028.1 GI:21760487

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Rukayotcha; Metacoa; Chordata; Craniata; Vertebrata; Euteleostomi;

1 (bases 1 to 640)

Wistow,G., Bernstein,S.L., Wyat,M.K., Ray,S., Behal,A.,

Touchman,J.W., Bouffard,G., Smith,D., and Peterson,K.

Expressed sequence tag analysis of human retina for the NEIRBank

Project: Retbinding, an abundant, novel retinal cDNA and alternative

splicing of other retina-preferred gene transcripts

Mol. Vis. 8 (4), 196-204 (2002)

22103461

12107411

COMMENT Contact: Wistow G

Section on Molecular Structure and Function

National Eye Institute

6/331, NIH, Bethesda, MD 20892-2740, USA

Tel: 301 402 3452

Fax: 301 496 0078

Email: graeme@helix.nih.gov

Plate: 03 row: d column: 11

Seq primer: M13RP1 reverse primer (ABI).

Location/Qualifiers

1..640

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="hd03d11"

/issue\_type="Retina"

/dev\_stage="Adult"

/lab\_host="EMDHI0B"

/clone\_lib="Human Retina cDNA (Un-normalized,

unamplified): hd/he"

/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue

was dissected from two 80 year old donors with no observed

eye disease. 100ug of total RNA was used for library

construction. A directionally cloned cDNA library in the

pSPORT vector (Life Technologies) was constructed at

Bioserve Biotechnology (Laurel MD) essentially following

the protocols of the Superscript Plasmid System full

details of which are contained in the manufacturer's

instruction manual (http://www.lifetech.com/). First

strand synthesis was carried out using a Not I

primer-adaptor

15'-pGACTGATCTTATGATCGAGCGCGCCCT(T)15-3'. EST analysis

was performed on the unamplified library at the NIH

Intramural Sequencing Center (NISC)."

# ORIGIN

## Alignment Scores:

Pred. No.: 5,416-117 Length: 640  
 Score: 1097.00 Matches: 209  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 88.47% Indels: 0  
 DB: 5 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x BQ636028 (1-640)

QY 31 ArgArgLeuSerLeuIleValSHISProGluCysSerThrMetSerGlyValAlaIle 50  
 DB 3 AGAAGACTTTCTTAATAAATCAATCCAGAGCTCCACCATGATGTTGGGAAAGCCATA 62  
 QY 51 GlnHisLeuAlaValGlnGlyValAsnArgPheHisAspIleValIleValSHISProPheLeuHis 70  
 DB 63 GAACATTTGGCCCAACAGAAATAGATTTCATTTTACATCAAACTCCCTTGACATAT 122  
 QY 71 AlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValIleThrAspLysIleIleMet 90  
 DB 123 GCTAAATATGTGATTTTCTTTACTGACATTCACACCGTTACTGATTAATAATATATG 182

QY 91 LysylsglyylsglygluylglyylleuylsglygluylleuylserSerAlaAlaaspIle 110  
 DB 183 AAAAAAGAAAAAGAGAGATATGAGAGGGCAATCTGCTTCAGAGAGACATT 242  
 QY 111 AlaAlaThValGlnHisThrMetAlaCysHisIleuValIysArgThrHisArgAlaIle 130  
 DB 243 GGTGGCAGCTACAGACACAAATGAGCATCTGTGTGAAAGACACATCGGGCTATT 302  
 QY 131 LeuPheCysIleGlnArgAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGly 150  
 DB 303 CTGTTTGTAGACAGAGAGACTGTTTACCTCAAAATATATGACATCTGCTTCATTCGCT 362  
 QY 151 GlyValAlaSerAsnBheTyrlleArgArgAlaIleuGlnIleuThrAsnAlaThrGln 170  
 DB 363 GGTGTGCAAGTAACCTTATATCCGAGAGCTCTGGAATTTTAAACAAAGCAACAG 422  
 QY 171 CysThrIleuLeuCysProProProlArgLeuCysThrAspAsnGlyIleMetIleAlaTTP 190  
 DB 423 TCCACTTGTGTGTCTCTCCAGACTATGACATGATATGATGATATGATGATGATG 482  
 QY 191 AsnGlyIleGlnArgLeuArgAlaGlyIleuGlnIleuThrAspIleGlyIleArg 210  
 DB 483 AATGGATTGAAAGACTACGCTGCTGCTGCTGCTTTCATGACATGAAAGGATCCGC 542  
 QY 211 TyrGluProLysCysProLeuGlyValAspIleSerIysGlyValGlyIleAlaSerIle 230  
 DB 543 TATGAACCAAAATGCTCTCTTGAAGTACATATCAAAAGAGTGAAGAGCTTCATA 602  
 QY 231 LysValProGlnIleuLysMetGluIle 239  
 DB 603 AAGTACCAATTAATAATGAGATA 629  
 RESULT 3  
 LOCUS AK045669 2284 bp mRNA linear HTC 03-APR-2004  
 DEFINITION Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length enriched library, clone:B230219017 product:similar to PUTATIVE SIALOGLYCOPROTEASE TYPE 2 (Homo sapiens), full insert sequence.  
 ACCESSION AK045669  
 VERSION AK045669.1 GI:26337528  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 AUTHORS Carninci, P. and Hayashizaki, Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE 99279253  
 PUBMED 10349636  
 REFERENCE  
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
 MEDLINE 20499374  
 PUBMED 11042159  
 REFERENCE  
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komno, H., Akiyama, J., Nishi, K., Kiteunui, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 TITLE RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer  
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
 MEDLINE 20530913  
 PUBMED 11076861

REFERENCE  
 AUTHORS 4  
 TITLE The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.  
 JOURNAL Functional annotation of a full-length mouse cDNA collection  
 REFERENCE Nature 409, 685-690 (2001)  
 REFERENCE  
 AUTHORS 5  
 TITLE The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
 JOURNAL Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 REFERENCE Nature 420, 563-573 (2002)  
 REFERENCE  
 AUTHORS 6 (bases 1 to 2284)  
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furum, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-JUL-2001) Yoshinori Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
 COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: http://genome.gsc.riken.jp/  
 URL: http://pantom.gsc.riken.jp/  
 URL: http://genome.gsc.riken.jp/  
 FEATURES  
 source  
 1..2284  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="PANTOM,DB:B230219017"  
 /db\_xref="taxon:10090"  
 /clone="B230219017"  
 /sex="male"  
 /issue\_type="corpora quadrigemina"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="adult"  
 221..1465  
 /note="unnamed protein product; putative similar to PUTATIVE SIALOGLYCOPROTEASE TYPE 2 (Homo sapiens) (SPTR|Q9H4B0, evidence: PASTY, 80%ID, 100%length, match=142)"  
 /protein\_id="BAC32450.1"  
 /codon\_start=1  
 /db\_xref="GI:26337529"  
 /translation="MIMRRPAGALIPKPKSKVGVGLRFSVHPRTLSCHKLVLGIETSCDRTGAAYDDEGVAGLALHSGQVHLKGTGIVPPAQQOLHRENIORIEYETLSACRITPSSLSAIVTIRKGLASLVGGLSLDVLVDFKPFPIHMEHALITRLTNKVEPFLVLLISGCHLVLGVGVDFLLGSLDAPMDIKVARRSLIKHPGSTMSGKALISGLAAGKATFPFTINPWNKNCDFSTGLQHTDKLITKKEBEGIKGQISSADIAAVOAHATFCHAKRTHRALIFCKOKNLSPAAVLVVGGSVSNLYIRKALIVANATQCTILCPPTCTDNGMIANNGIRLRAGLGVLDVDIRPBCPLGVDSISREVAERAIKPRKML"  
 2262..2267  
 /note="putative"  
 2284  
 /note="putative"  
 ORIGIN  
 polyA\_signal  
 polyA\_site



```
source
1..1622
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM.DB:2610001M19"
/db_xref="taxon:10090"
/clone="2610001M19"
/issue_type="whole body"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days embryo"
207..1451
/note="unnamed protein product; putative
similar to PUTATIVE SIALOGLYCOPROTEASE TYPE 2 [Homo
sapiens] (SPTR|Q9H4B0, evidence: FASTY, 80%ID, 100%length,
match=1242)"
/protein_id="BA027506.1"
/db_xref="GI:12847276"
/translation="MLMRTAGAIKPKPKYVGLRFRSHPRTLSCKLVIGIET
SCDDTGAADVETGVIGLALHSGVLSQVHLTGIVPPVAQQLRENIQRIVEETLSAC
RTSPDLASIAITIKPGLALSGVLSFSLQVNPCKPPIPHHMAHALITRTK
VPPPLVILISGHICLALVGVSDPLILGKSLDIAKGMVLDVARRSLIKPECT
MSGKAIKQAKGNRPHFTINPPMNAKCDSPFGLOHTTKLITKKEGIEKE
QILSSADIAAAVOHATACHLARTHRALIFCKQKILSPANVLVSGVASNLVLR
KALEIVANAQCTLCPPRLCTDNGIMIAMNGIERLRAGLGVLDVEDIRPEKPL
GVDSREVAEAAIKVPRLMKML"
1605..1610
/note="putative"
1622
/note="putative"
1622
/note="putative"

polya_signal
polya_site

ORIGIN
Alignment Scores:
Pred. No.: 2.16e-113 Length: 1622
Score: 1071.00 Matches: 204
Percent Similarity: 91.21% Conservative: 14
Best Local Similarity: 85.36% Mismatches: 21
Query Match: 86.37% Indels: 0
DB: Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x AK011265 (1-1622)
OY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlnGlySerLeuAspIle 20
DB 732 CTTTGGCATTAGTCCAGAGGTTGTCGATTTCTGCTCTTGGCAAGCTCTTTGACATA 791
OY 21 AlaProGlyAspMetLeuAspIleValAlaArgArgLeuSerLeuIleValHisProGlu 40
DB 792 GACCCAGGGCAGACGCTTGACAGAGTGCGAAGAAGACTTTCTTAATCAACATCCAGAA 851
OY 41 CysSerThrMetSerGlyGlyValAlaIleGlnHisLeuAlaValGlnGlyAsnArgPhe 60
DB 852 TGTTTTACAAATGAGTGTGAGAAAAGCTATAGAACAGTTGGCCAAAGACGGAATAGATTTC 911
OY 61 HisPheAspIleLeuSerProLeuHisHisAlaValAsnCysAspPheSerPheThrGly 80
DB 912 CATTTTACTATCAATCCACCTATGACAGAAATGCTAAGAAATGCGATTTTCTTACCGGGA 971
OY 81 LeuGlnHisValThrAspIleIleMetLeuValGlnGlyGlnGlnGlyIleGlnGly 100
DB 972 CTTCAACATTTTCTGATTAAGCTTAATACACACAGAAAAGAAAAGAGGCAATTGAGAG 1031
OY 101 GlnGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120
DB 1032 GAGCAATCTCGTCATCAGCTGACAGACATTGCTGCGGTACGACATGCAACGCGTGC 1091
OY 121 HisLeuValIysArgThrHisAlaGalaIleLeuPheCysIysGlnArgAspLeuLeuPro 140
DB 1092 CACCTTGCGAAAAGAACACATCCGCGTATCTGTTTTCAGAGCAAGAAAATTTCTCTCT 1151
OY 141 GlnAsnAspAlaValLeuValAlaSerGlyValAlaValAsnAspPheThrIleArgArg 160
DB 1152 CCAAGTAACGACAGTATGTTGATTCGAGGTTGCAAGTAACTTGTAATACCGCAAAA 1211
```

```
OY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrIleLeuLeuCysProProProArgLeu 180
DB 1212 GCATTGAAATTTGTCGCAAAATGCAACGACGACGCTGTTGTGTGTCACCTCCCAAGACTG 1271
OY 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGlnIleValArgAlaGlyLeu 200
DB 1272 TGCACTACCAATGGCATCATGATATGCAATGGAATTAAGATTACGTTGCGGCTTG 1331
OY 201 GlyIleLeuHisAspIleGlnGlyIleArgGlyArgProLysCysProLeuGlyValAsp 220
DB 1332 GCGCTTTACATGATGTAGGAAGACATCCGATATGACCAAAATGCTCTTGAGTAAAC 1391
OY 221 IleSerIysGlnValGlnGlnAlaSerIleValValProGlnLeuIleMetGlnIle 239
DB 1392 ATATCCAGAGAAATTGCAAGAGCTGCATTAAGATGACGCGATTAAGATGCACTT 1448

RESULT 5
BX391919/c 852 bp mRNA linear EST 28-APR-2004
LOCUS BX391919 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
DEFINITION cDNA clone CS0DK001YE02 3-PRIME, mRNA sequence.
ACCESSION BX391919
VERSION BX391919.2 GI:46846154
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 852)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLES Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On May 13, 2003 this sequence version replaced gi:30611736.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime
end enriched, double-strand cDNA was digested with NotI and cloned
into the NotI and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
1240.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?b=CS0BA1035ZEB07_CS03317_1&c=1240.r

FEATURES
source
1..852
/location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK001YE02"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with NotI and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized."

ORIGIN
Alignment Scores:
Pred. No.: 2.28e-97 Length: 852
Score: 930.00 Matches: 180
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 75.00% Indels: 0
DB: Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x BX391919 (1-852)
OY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlnGlySerLeuAspIle 20
```

Db 542 CTGTTGGCATTAGTAAAGAGATTTCAGATTTCTCTTCTTGGAAAGCTTTGGACATA 483  
Oy 21 AAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
Db 482 GCACCAAGGACATGCTTATCAACAGGTGGCAAGAGACTTTCTTATTAACATCCAGAG 423  
Oy 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
Db 422 TCTCCACCATGATGATGGTGGAAAGCCATAGAACATTGGCCAAACAGAAATAGATT 363  
Oy 61 HisPheAspIleLysProPheLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80  
Db 362 CATTGTGACATCAAACTCCCTGTCATCAAGCTAAAAATTTGATATTTCTTTTACGGA 303  
Oy 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGluGlyIleGluLys 100  
Db 302 CTTCACACAGTTTACTATTAATTAATATCAAAAGAGAAAGAGAGATTGAGAGAG 243  
Oy 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120  
Db 242 GGGCAAAATCTGTTCTTACAGACAGACATTGGTCCACACTACAGACACAAATGGCATGT 183  
Oy 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
Db 182 CATCTTGAAAAAGAACACATCGGGCTATTCTGTTTGTAAAGACAGAGACTGTACT 123  
Oy 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTrpIleArgArg 160  
Db 122 CAAAATAAGTCAGATCTGTTGATCTGTTGATCTGTTGATCTGTTGATCTGTTGATCTG 63  
Oy 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
Db 62 GCTCTGGAATTTTAAACAACGACACAGTGCATTTGTTGTCTCTCTCCACACTA 3

RESULT 6  
BC030671 1609 bp mRNA linear HTC 19-NOV-2003  
LOCUS Mus musculus O-sialylglycoprotein endopeptidase-like 1, mRNA (CDNA  
DEFINITION clone IMAGE:1226118), containing frame-shift errors.  
ACCESSION BC030671  
VERSION BC030671.1 GI:21040459  
KEYWORDS HTC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 1609)  
AUTHORS Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D.,  
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.L., Wang, J., Hsieh, F.,  
Diatchenko, L., Marusina, K., Farmer, A.G., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Scharf, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
Carinci, P., Prange, C., Raha, S.S., Loggiano, N.A., Peters, G.J.,  
Abrahamson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huylk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schultz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.E.,  
Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.,  
human and mouse cDNA sequences of more than 15,000 full-length  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
MEDLINE 22388257  
PUBMED 12477932  
REFERENCE 2 (bases 1 to 1609)  
AUTHORS Strausberg, R.

TITLE Direct Submission  
JOURNAL Submitted (20-MAY-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Marcello Bento Soares, Ph.D.  
cDNA Library Preparation: M. Bento Soares, University of Iowa  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
contact: [amadad@systemsbiology.org](mailto:amadad@systemsbiology.org)  
Anup Madan, Jessica Fahy, Erin Helton, Mark Kettelman, Anuradha  
Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting

FEATURES  
source  
1. 1609  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1226118"  
/tissue\_type="Thymus gland, mouse"  
/clone\_id="Soares\_thymus\_2NBMT"  
/lab\_host="DH10B"  
/note="Vector: pT73-Pac"

ORIGIN  
Alignment Scores:  
Pred. No.: 1,656-96 Length: 1609  
Score: 926.00 Matches: 183  
Percent Similarity: 81.59% Conservative: 12  
Best Local Similarity: 76.57% Mismatches: 20  
Query Match: 74.68% Indels: 24  
DB: Gaps: 1

US-10-649-273-2\_copy\_176\_414 (1-239) x BC030671 (1-1609)  
Oy 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
Db 749 CTGTTGGCATTAGTCAAGGATTTCCGATTTCTGCTTGGGAAGTCTTTGGACATA 808  
Oy 21 AAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
Db 809 GCACCAAGGACATGCTTATCAACAGGTGGCAAGAGACTTTCTTATTAACATCCAGAA 868  
Oy 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
Db 422 TCTCCACCATGATGATGGTGGAAAGCCATAGAACATTGGCCAAACAGAAATAGATT 928  
Oy 61 HisPheAspIleLysProPheLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80  
Db 929 CATTTCATCAATTCACCATGATGATGATGATGATGATGATGATGATGATGATGATG 988  
Oy 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGluGlyIleGluLys 100  
Db 989 CTTCACACAGTTTACTATTAATTAATATCAAAAGAGAAAGAGAGATTGAGAGAG 1048  
Oy 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120  
Db 1049 GGGCAAAATCTGTTCTTACAGACAGACATTGGTCCACACTACAGACACAAATGGCATGT 1108  
Oy 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
Db 1109 CACCTTGGAATAAACAACATCGGGCTATTCTGTTTTCAGACAGAAATTTGCTCTCT 1168

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAspHeterYrileargArg 160  
 DB 1169 CCAGCTAACGAGATTAAGTTATCTGAGAGGTTGCAAGTAACTGTATCATCCGAAA 1228  
 QY 161 AlaLeuGluIleLeuThraAsnAlaThrGlnCysThreLeuGlyCysProProPheArgLeu 180  
 DB 1229 GATTGGAAATGTTCCCAATGCAAGCAGTGCAGCTGTGTGTGTCCACTCCCAAGACTG 1288  
 QY 181 CysThraAsnGlyIleLeuIleAlaIlePheAsnGlyIleGluArgLeuAlaGlyLeu 200  
 DB 1289 TGCACGACAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 1315  
 QY 201 GlyIleLeuHisAspIleGluGlyIleArgGlyIleProLeuGlyCysProLeuGlyValAlaAsp 220  
 DB 1316 -----TGAATCTCTCTTGGAGTAC 1336  
 QY 221 IleserLeuGluValGlyGluAlaSerIleLeuValProGlnLeuGlyMetGluIle 239  
 DB 1337 AATTCAGAGAGAGTGCAGAGAGTGCAGAGAGTGCAGAGAGTGCAGAGAGTGCAGAG 1393  
 RESULT 7  
 BE740611 701 bp mRNA linear EST 15-SEP-2000  
 LOCUS 601595739F1 NIH\_MGC\_9 Homo sapiens CDNA clone IMAGE:3949640 5'  
 DEFINITION mRNA sequence.  
 ACCESSION BE740611 GI:10154603  
 VERSION BE740611.1  
 KEYWORDS Homo sapiens (human)  
 SOURCE EST.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE NIH-MGC http://mgi.nci.nih.gov/  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: cgabs-remail.nih.gov  
 Tissue Procurement: DCTD/DTF  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMLU)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LMLU at: image.llnl.gov  
 Plate: L1C8614 row: n column: 09  
 High quality sequence stop: 701.  
 Location/Qualifiers  
 1..701  
 source  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3949640"  
 /tissue\_type="adenocarcinoma cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_9"  
 /note="Organ: ovary; Vector: pOT87; Site 1: XhoI; Site 2:  
 EcoRI; CDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCACGAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-CDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 6,31e-95 Length: 701  
 Score: 908.00 Matches: 178  
 Percent Similarity: 98.91% Conservative: 3  
 Best Local Similarity: 97.27% Mismatches: 0  
 Query Match: 73.23% Indels: 2  
 DB: 2 Gaps: 0

US-10-649-273-2\_copy\_176\_414 (1-239) x BE740611 (1-701)  
 QY 1 LeuLeuAlaLeuValGlnGlyValSerAspHeterLeuLeuGlySerLeuAspIle 20  
 DB 155 CTGTGGCATTAAGTTCAAGAGGTTTCAGATTTTCTGCTTCTTGAAGATCTTTGGACATA 214  
 QY 21 AlaProGlyAspMetLeuAspPheValAlaArgArgLeuSerLeuIleHisProGlu 40  
 DB 215 GCACAGAGTGACATGCTTGAAGAGTGCAAGAGACTTTCTTATATAAATCATCCAGAG 274  
 QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaGlyGlnGlyAsnArgPhe 60  
 DB 275 TGCTCCACCAAGAGTGGGAGAAAGCCATMAACATTTGGCCAAACAGAAATGATTT 334  
 QY 61 HisPheAspIleLeuProProLeuHisAlaIleAsnGlyAspPheSerPheThrGly 80  
 DB 335 CATTTTGACATCAACCTCCCTTGATCATCTCTAATAATGTGATTTTCTTTTACTGGA 394  
 QY 81 LeuGlnHisValIleThraAspPheIleIleMetLys--LysGluLeuGluGluGlyIleGlu 100  
 DB 395 CTTCAACAGCTTACTGATTAATAATATATGAAAACAGAAACAGAGAGATTTGAGA 454  
 QY 100 YsgIyGlnIleLeuSerSerAlaAlaAspIleAlaIleValGlnHisThrMetAlaC 120  
 DB 455 AGGGGCAAAATCTGCTTTCAGACAGACAGATTTGCTGCACAGTACAGACAAATGGCAT 514  
 QY 120 YshIleLeuValIleArgThrHisArgAlaIleLeuPheCysLeuGlnArgAspLeuLeuP 140  
 DB 515 GTCAATCTTTGAAAAAGAACACATCGGGCTATTTGTTTGAAGAGAGAGCTTTGATC 574  
 QY 140 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspHeterYrileargArg 160  
 DB 575 CTCAAAATTAAGCACTGCTGTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 634  
 QY 160 GlnAlaLeuGluIleLeuThraAsnAlaThrGlnCysThreLeuGlyCysProProPheArg 180  
 DB 635 GAGCTGTGAATTTTAAACAACGACACAGTGACCTTTGTGTGTCTCTCCCAAGC 694  
 QY 180 euGys 181  
 DB 695 TATGC 699  
 RESULT 8  
 CF362328/c 658 bp mRNA linear EST 25-AUG-2003  
 LOCUS 829596 MARC 3P1G Sus scrofa CDNA 3', mRNA sequence.  
 DEFINITION CF362328  
 ACCESSION CF362328.1 GI:34161882  
 VERSION EST.  
 KEYWORDS Sus scrofa (pig)  
 SOURCE Sus scrofa  
 ORGANISM Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
 REFERENCE Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Fox,J.,  
 Wise,T.A., Noneman,D.J., Wray,J.E. and Keele,J.W.  
 A second set of porcine ESTs from a pooled-tissue normalized  
 library  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@email.marc.usda.gov  
 Single pass sequencing. Bases called with phred v0.020425.c and  
 trimmed with the aid of the trim\_alt option. Vector identified with  
 cross\_match v0.990329.  
 Plate: SRG8015 row: H column: 12  
 Seq primer: TAGAAGCAGCAGTCAGG.  
 FEATURES  
 source  
 1..658  
 Location/Qualifiers

/organism="Sus scrofa"  
/mol\_type="mRNA"  
/db\_xref="taxon:9823"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/clone\_lib="MARC 3Pig"  
/note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;  
Library made with RNA pooled from multiple tissues  
including brain, liver, muscle, placenta/endometrium,  
ovary, testes, and bone marrow."

## ORIGIN

Alignment Scores:  
Pred. No.: 3,07e-91 Length: 658  
Score: 876.00 Matches: 165  
Percent Similarity: 93.68% Conservative: 13  
Best Local Similarity: 86.84% Mismatches: 12  
Query Match: 70.65% Indels: 0  
DB: 7 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x CF362328 (1-658)

```

OY      1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlnGlySerLeuAspIle 20
DB      630 CTTTGGCATTGATTAGAGAGATTTCAGATTTCCTCTTGACAGCTTTGGACATA 571
OY      21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB      570 GCACCAAGGTGACATGCTTACACAGAGTACAGAGACCTTTCTTAATAAACATCCAGAG 511
OY      41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
DB      510 TGCTCCACCATGATGCTGTGGAGAGCCATTAAGACATTTCGCAACAGGAATAAGTTG 451
OY      61 HisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80
DB      450 CATTGTGATTTCAAACTCCCACTGATTAATAAATTGATTTTCTTTTCTGGA 391
OY      81 LeuGlnHisValThrAspLysIleIleMetLysGlyGlnGlyGluGlyIleGluLys 100
DB      390 CTTCAACATGATTATTATTAAGACATTAATGCAAGAGAAAAGAGAGGATTGAGAGAG 331
OY      101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB      330 GGGCAAAATCTGCTCTTCACTGACAGCATTTGCTGCTGCACTACACACAGTAGCTTGC 271
OY      121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB      270 CATATTGCAAAAAGAACTCATCTGCTATTCTTTTGCAGAAAGAGAGACTTATTATGT 211
OY      141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrrIleArgArg 160
DB      210 CAAAGTAATGCAATATTGTTGATCTGAGAGGTTGCAAGTAATTATTAATACGAAA 151
OY      161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
DB      150 GCTTTGAAGCTGTGCAATGCAACATGCACTCTGTGTCTCTCTCCAGACTA 91
OY      181 CysThrAspAsnGlyIleMetIleAlaTyr 190
DB      90 TGCACGTGATTAATGCAATTAATGATTCATGG 61

```

## RESULT 9

CK941819/c CK941819 637 bp mRNA linear EST 15-MAR-2004  
LOCUS 4065407 BARC 10BOV Bos taurus cDNA clone 10BOV12\_F24 3', mRNA  
DEFINITION sequence.

ACCESSION CK941819 GI:45456199  
VERSION CK941819.1  
KEYWORDS EST.  
SOURCE Bos taurus (cow)  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovinae; Bos.  
1 (bases 1 to 637)  
Sonstegard, T.S., Van Tassel, C.P., Matukumalli, L.K., Hartney,  
G.P., Bosak, S., Rubenfield, M. and Gabarre, L.C.  
Production of EST from cDNA libraries derived from immunologically  
activated bovine gut  
Unpublished (2004)  
Contact: Tad S. Sonstegard  
Bovine Functional Genomics Laboratory  
Animal and Natural Resources Institute  
B19, 200 Km2A BARC-Bast, Beltsville, MD 20705, USA  
Tel: 3015048416  
Fax: 3015048414  
Email: tads@anri.barc.usda.gov

Single pass sequencing. Bases called and trimmed with phred  
0.000925 using options -trim alt -trim faste. Vector identified  
by cross\_match using options -mismatch 12 -minscore 18  
Plate: 12 row: F column: 24  
Seq primer: AGCGATACACATTTCACACAGG  
High quality sequence stop: 637.

## FEATURES

## source

1..637  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/strain="Holstein"  
/db\_xref="taxon:9913"  
/clone="10BOV12\_F24"  
/sex="male"  
/tissue\_type="pooled"  
/dev\_stage="Multiple"  
/lab\_host="DH10B T1 phage resistant"  
/clone\_lib="BARC 10BOV"  
/note="Organ: Small Intestine; Vector: pagen-1; Site 1:  
EcoRV; Site 2: NotI; Equimolar amounts of mRNA extracted  
from proximal jejunums of 18 and 21 wk old steers, and  
distal ileums of 14 day old calves. proximal jejunum  
exposed to C. oncophora for 3 and 6 weeks, and distal  
ileum exposed to C. parvum for 7 days"

## ORIGIN

Alignment Scores:  
Pred. No.: 7.81e-87 Length: 637  
Score: 838.00 Matches: 153  
Percent Similarity: 92.47% Conservative: 19  
Best Local Similarity: 82.26% Mismatches: 14  
Query Match: 67.58% Indels: 0  
DB: 7 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x CK941819 (1-637)

```

OY      54 AlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLysAsn 73
DB      632 GCCAAACAGGAAATATATGATTCATTTCAGGCTCCCAACGTCGTAATAAT 573
OY      74 CysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysGln 93
DB      572 TGTGATTTCTTTTCTGACCTTCAACACCTTAATTGAATGATTAATCAAAAGGA 513
OY      94 LysGlnGlnGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThr 113
DB      512 AAAGAGAGAGTATCGAGAGGCGGCAAGTCTGCTTCACTGACAGCATTTGCTTGGC 453
OY      114 ValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCys 133
DB      452 GTGACAGCACACCGGCTCGCAATTCGCAAAAAGAACACATCGTCTCTTCTGTCTGC 393
OY      134 LysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAla 153
DB      392 AAGCAGAGAGGCTTCTTAAGTCAGAGTACGACATCTGTTGATTCGAGGCTGCCA 333
OY      154 SerAsnPheTyrrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeu 173

```



Db 332 AGTAATTATATATCGAAAACCCCTGGAATTTGTGACCAATGACACAGTCGACTTTG 273  
 QY 174 LeuCyProProProArgLeuCyThrAspAsnGlyIleMetIleAlaTPAsnGlyIle 193  
 Db 272 CTGTGCGCGCCGCCAGACTTTGCACTGACCAACGGCTTATGATTTGATGAAATGGTGT 213  
 QY 194 GUAATGLeuAAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGluPro 213  
 Db 212 GAAAGACTACGTGCTGCTGGCTGACATTTTACACACACACAGAGCATCCGCTACGAACCA 153  
 QY 214 LysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysValPro 233  
 Db 152 AAATGTCCTCTTGAGATGATATATATCAAAAAGAGTTGAGAGAGCTGCTATTAATAAGTCCA 93  
 QY 234 GlnLeuLysMetGluIle 239  
 Db 92 AGATTAAATGAGATT 75  
 RESULT 10  
 BQ961028 922 bp mRNA linear EST 21-AUG-2002  
 LOCUS AGENCOURT\_8863711 NIH\_MGC\_18 Homo sapiens cDNA clone IMAGE:6423902  
 DEFINITION 5', mRNA sequence.  
 ACCESSION BQ961028.1 GI:22376506  
 VERSION BQ961028  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 922)  
 AUTHORS NIH-MGC http://imgc.ncbi.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaabds-remail.nih.gov  
 Tissue Procurement: DCTD/DTP/Gazdar  
 cDNA library Preparation: Rubin Laboratory  
 cDNA library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: l1CM2603 row: d column: 15  
 High quality sequence stop: 584.  
 Location/Qualifiers  
 1..922  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6423902"  
 /tissue\_type="large cell carcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 18"  
 /note="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into BcoRI/XhoI sites using the following 5' adaptor:  
 GGCAACAG(G). Library constructed by Ling Hong in the  
 laboratory of Gerald M. Rubin (University of California,  
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
 Superscript II RT (Life Technologies). Note: this is a  
 NIH\_MGC library."

## ORIGIN

## Alignment Scores:

Pred. No.: 4.6e-82 Length: 922  
 Score: 799.00 Matches: 179  
 Percent Similarity: 88.04% Conservatives: 5  
 Best Local Similarity: 85.65% Mismatches: 16  
 Query Match: 64.44% Indels: 10  
 DB: 5 Gaps: 3

US-10-649-273-2\_COPY\_176\_414 (1-239) x BQ961028 (1-922)

QY 1 LeuLeuAlaLeuValGlnGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
 Db 291 CTGTGGCACTATTAGTTCAAGGAGTTTCAGATTTTCTGCTTCTTGAAAATCTTTGACATA 350  
 QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
 Db 351 GCACCAAGTGCAGCTTGACAGAGTGCCAGAAAGACTTTCTTTATATTAACATCCAGAG 410  
 QY 41 Cys-SerThrMetSerGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPh 60  
 Db 411 TGCATCCACCATGAGTGTGGGAAAGCATATGAACATTGGCCAAACAGGAATAGATT 470  
 QY 60 eHisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGl 80  
 Db 471 TCATTTGACATCAACCTCCCTTGATCATGTAATAAATTTGATTTTCTTTACTCG 530  
 QY 80 YLeuGlnHisValThrAspLysIleIleMetLysLeuLysGluGluGlyIleGluLys 100  
 Db 531 ACTTCAACACCTTACTGATTAATAATATATATAAAGAAAGAGAGATTATGAGAA 590  
 QY 100 SGLYGLIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCy 120  
 Db 591 AGGCAAAATCCTGCTTTCAGACGACGACATTCCTGCCACATGACGACACATGGCATG 650  
 QY 120 SHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPr 140  
 Db 651 TCATCTGTGTAAGAAAGAA-CATCGGGCTATTCTGTTTGTAAAGAGAGACTTGATACC 709  
 QY 140 OGLAsnAsnAlaValLeuValAlaSer-GlyGlyValAlaSerAsnPheTyrIleArgA 160  
 Db 710 TCAAAATATATCACTACTGCTGATCTGCGGGGGGTGCAAGTAACTTATATATCCCGC 769  
 QY 160 rGAlaLeuGlnIleLeuThr-AsnAlaThrGlnCysThr-LeuLeuCyProProArg 179  
 Db 770 AAACCTCGAAAATTTTACAAACCAACAGGGGACCTTGTGTGGTCCCTCCCA 829  
 QY 179 gLeuCyThr-AspAsnGlyIleMetIleAlaTTP-----AsnGlyIleGluArg--L 196  
 Db 830 ACTATGCACTGATATATGG---CATTTATGATGCTGCGGAGGAGATTGAAAAAATA 886  
 QY 196 euArgAlaGlyLeuGly 201  
 Db 887 CMTGCTGCGCTTGGGG 903

## FEATURES

## source

1..922  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6423902"  
 /tissue\_type="large cell carcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 18"  
 /note="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into BcoRI/XhoI sites using the following 5' adaptor:  
 GGCAACAG(G). Library constructed by Ling Hong in the  
 laboratory of Gerald M. Rubin (University of California,  
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
 Superscript II RT (Life Technologies). Note: this is a  
 NIH\_MGC library."

RESULT 11  
 AV602901/c 545 bp mRNA linear EST 27-NOV-2001  
 LOCUS AV602901 Bos taurus kidney fetus Bos taurus cDNA clone EIK1013A07  
 DEFINITION 3', mRNA sequence.  
 ACCESSION AV602901  
 VERSION AV602901.1 GI:9725227  
 KEYWORDS EST.  
 SOURCE Bos taurus (cow)  
 ORGANISM Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovinae; Bos.  
 REFERENCE 1 (bases 1 to 545)  
 AUTHORS Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.  
 TITLE Establishment of a high throughput EST sequencing system using  
 poly(A) tail-removed cDNA libraries and determination of 36,000  
 bovine ESTs  
 Nucleic Acids Res. 29 (22), E108 (2001)  
 JOURNAL MEDLINE 21570554  
 PUBMED 11713328  
 COMMENT Contact: Yoshikazu Sugimoto  
 Animal Genetics Division  
 Shitakawa Institute of Animal Genetics  
 Odakura, Nishigo, Nishi-shitakawa, Fukushima 961-8061, Japan  
 Tel: 81-248-25-5641  
 Fax: 81-248-25-5725



Db 220 GCTATTCTGTTTGTAGACAGAGACTTGTACTCTCAAAAATATGACGACTGTTCCG 279  
Qy 149 SerGIgLVAlAlaserapheryllleargArAlaleuGIuileuthrasnAla 168  
Db 280 TTTGGTGGTGGCAAGTAACTTCTATATCCGACAGCTCTGGAATTTTAAACAAGCA 339  
Qy 169 ThrgInCyethrleuLeuCySproProArXleuCySthrasphangilyllemetile 188  
Db 340 ACACAGTGCATCTTGTGTCTCTCCACAGACTATGACATGATATGACATATGATT 399  
Qy 189 AlAtPAserGIyIlleGIuArGIeAlaGIleuGIyIlleuHIsapIlleGIuGIy 208  
Db 400 GGATGGAAAGTATTAAAGACTACGCTGCTGGCATTTTACATGACATAGAAGGC 459  
Qy 209 lIeArGIyGIuProIyCySPro 216  
Db 460 ATCCGATGAACCAAAATGTCCTC 483

RESULT 13  
CF257246 822 bp mRNA linear EST 07-AUG-2003  
LOCUS phn008\_902 PHA-activated splenocytes Gallus gallus cDNA, mRNA  
DEFINITION sequence.  
ACCESSION CF257246  
VERSION CF257246.1 GI:33490501  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 822)  
Witzell,H., Bed'Hom,B., Morin,V., Young,J.R., Whitaker,C.J.,  
Chausse,A.M. and Zoorob,R.  
A collection of chicken ESTs from activated immune cells  
Unpublished (2003)  
CONTACT: Zoorob R  
COMMENT UPR 1983

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
UPR 1983

FEATURES  
source  
1..822  
Location/Qualifiers  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9031"  
/cell\_type="splenocytes"  
/clone\_lib="PHA-activated splenocytes"  
/note="Vector: pTriblEX2"

ORIGIN  
Alignment Scores:  
Pred. No.: 2.85e-80 Length: 822  
Score: 783.00 Matches: 150  
Percent Similarity: 79.57% Conservative: 37  
Best Local Similarity: 63.83% Mismatches: 48  
Query Match: 63.15% Indels: 1  
DB: 7 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x CF257246 (1-822)

Qy 1 leuLeuAlaleuAlaInglyValSerAspPheLeuLeuGIySerLeuAspIle 20  
Db 119 ATCTTGACATGACACAGAGAGTTTCAGATTCTCTGCTGACACTCATATGATACA 178  
Qy 21 AlAProGIyAspMetLeuAspLyValAlaArgArgLeuSerLeuIleYSHISProGIu 40  
Db 179 GACUCAGTACATGTTGGATAGTACAGAAAGGCTCTTTAGTGAACACCCGAG 238  
Qy 41 CySerThreSerGIyGIyValAlaIleGIuHIsleuAlaYsGIuGIyAsnArgPhe 60

Db 239 TGCCACGACATGCCCGGGGGAGAGCAATAGACACCTGGCTCAAAACCGAGACTGGCA 298  
Qy 61 HisPheAspIleYsProProLeuHIsAlaYsAenCyAspPheSerPheThrgly 80  
Db 299 CAGTACACTTCAAGACTTCCCAACAGATGATGACTGATTTCTTCTCCCGCA 358  
Qy 81 LeuGIuHIsAlaThrAspLySIlleIleMetYsGIuYsGIuGIuGIyIlleGIuYs 100  
Db 359 CTTCAGAGCCCTTCACAAAGCCATCTTCAAGAAAGAAAGAAAGAAAGATTCACAA 418  
Qy 101 GlyGIuIlleuSerSerAlaAlaAspIleAlaIleThrValGlnHIsThreAlaCyS 120  
Db 419 GGGGAATTCCTGTCCTCGCTTAAGACATCCCTGCTGTCGACAGACTAGTGTCTCT 478  
Qy 121 HisLeuValYsArgThrHisArgAlaIleLeuPheCySlySgInArgAspLeuPro 140  
Db 479 CATATTTCAGCGGACACACCGACCATGCTCTTCTGCATGAAGAAACAGATATTATTA 538  
Qy 141 GlnAsnAsnAlaValleuValAlaSerGIyGIyValAlaSerAspPheYrIlleArg 160  
Db 539 CCAAAACTGCACACTGCTGTGTATCAGAGAGAGTGCAGATATACATATATCAGAAA 598  
Qy 161 AlaleuGIuileuthrasnAlaThrgInCySthrleuLeuCySproProArXleu 180  
Db 599 GGACTGCAGACTCTGGCAAAATGCAACGGTTTCTGCTCTCTCTCCCAAGGCTG 658  
Qy 181 CyThrAspAsnGIyIlleMetIlleAlaThrAsnGIyIlleGIuArgLeuAlaGIyLeu 200  
Db 659 TGCAACGATATGATGTTATATTCATGATGATGATGATGATGATGATGATGATGAT 718  
Qy 201 GIYIleuHIsAspIleGIyIlleArgYrGIuProIyCySProLeuGIyValAsp 220  
Db 719 GGTATTTTATACAGTACATGATGATGATGATGATGATGATGATGATGATGATGAT 778  
Qy 221 lIeSerlySgIuValGIyGIuAlaSerIleYsValProGlnLeu 235  
Db 779 ATTCCAAAGAGTTGAA-GAGGATTCATCAAGTCCCAAGCTCA 822

RESULT 14  
BU127463 866 bp mRNA linear EST 25-NOV-2002  
LOCUS 603114407F1 CSECHU20 Gallus gallus cDNA clone Chest6m2 5', mRNA  
DEFINITION sequence.  
ACCESSION BU127463  
VERSION BU127463.1 GI:25338728  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 866)  
Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,  
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.  
A Comprehensive Collection of Chicken cDNAs  
Curr. Biol. 12 (22), 1965-1969 (2002)  
22335534  
12445392  
Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)  
PO Box 88, Manchester, M60 10D, UK  
Tel: 01612008930  
Fax: 01612360409  
Email: Simon.Hubbard@umist.ac.uk.

FEATURES  
source  
1..866  
Location/Qualifiers  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="White Leghorn, HiseX"  
/db\_xref="taxon:9031"  
/clone="CHEST6m2"

```

/dev stage="36"
/lab_host="DH108"
/clone_lib="CSECHL20"
/notes="Organ: limbs; Vector: pBluescript II KS(+); Site 1:
EcoRI; Site 2: NotI; Modification of pBluescript II KS(+)
[is]tagene] vector to accommodate cDNA produced with the
T-primed protocol (construction of uni-directionally
cloned cDNA libraries from messenger RNA for improved 3'
end DNA sequencing by Glenn Py, et al. U.S. Patent #
6,387,624). Cut pBluescript II KS(+) with NotI and EcoRI.
ligate in double stranded adaptor containing BspGI and
BamHI sites [5'ggccggcgccgagcccgagtcgcaaaaaag]
[5'aattcttttccggatccggggcgccgagc]"

```

## ORIGIN

## Alignment Scores:

```

Pred. No.: 9,58e-77 Length: 866
Score: 753.00 Matches: 149
Percent Similarity: 78.81% Conservative: 37
Best Local Similarity: 63.14% Mismatches: 50
Query Match: 60.73% Indels: 2
DB: 5 Gaps: 0

```

US-10-649-273-2\_COPY\_176\_414 (1-239) x BUI27463 (1-866)

```

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlnGlySerLeuAspIle 20
DB 25 ACTCTGGCAGTAGACACAGAGAGTTTCAGATTCTTCTGCTGGACAGTCATAGATATA 84
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysIleProGlu 40
DB 85 GCACCGAGGTGACATGCTGATAGTAGCAAGAGGCTCTTCTTATGGAAGCACCCGAG 144
QY 41 CysSerThrMetSerGlnGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 145 TCCACCGGATGGCTGGGGGAAAGCAATAGACAGCTGCTCAACCGGAGACTGGCAA 204
QY 61 HisPheAspIleLysProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80
DB 205 CAGTACACTTTCAGATCTCCCATGCAACAGTACGTACGTATTTCTTCTCCGGA 264
QY 81 LeuGlnHisValThrAspLysIleIleMetLysGlnGlyGlnGlyIleGlnLys 100
DB 265 CTTCAGAGGCTGTCTGCAAAAGCCATTCTTCAGAAAGAAAGAAAGGTAATTCAGAA 324
QY 101 GlnGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB 325 GGGGAAATCTGCTGCGCTTAAGACATCGCTGCTGCACAGCATGTAGTGGCTGCT 384
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140
DB 385 CATATTATTCAGCGGACACACAGCCAGCTCTTCTGCATGAAAAACACATATTATTA 444
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaLaserAsnPheryIleArgArg 160
DB 445 CCAAAAACAGCAACTGCTGGTGTATCAGAGGAGGTTCAGATATCATATACAGAAA 504
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
DB 505 GGACTGACAGACTCTGGC-AATGCAACGGTTTGTCTTTCTGCTCTCCCAAGGCTG 563
QY 181 CysThrAspAsnGlyIleMetIleAlaTPAsnGlyIleGlnArgLeuAArgAlaGlyLeu 200
DB 564 TCCACCGGATAGTGTATGATTTGATGAGAAAGGCAATTAAGAGTTGCGGAGGAGTGT 623
QY 201 GlnIleLeuHisAspIleGlnGlyIleArgGlyArgLysProLysCysProLeuGlyValAsp 220
DB 624 GGTATATTA-TACAGTACTATATGCAATCCGCTATGAAACCAAGCTCCCTTGGATTTGAT 682
QY 221 HisSerLysGlnValGlnLysValSerIleLysValProGlnLeuLys 236
DB 683 ATTTCAAAAGAGTTGAAGGAGATTCATCAGAGTCCAGACTAAAG 730

```

## RESULT 15

CK365185 736 bp mRNA linear EST 23-DEC-2003

## LOCUS

CK365185

## DEFINITION

AGENCOURT 17157291 NIH MGC 231 Rattus norvegicus cDNA clone

## ACCESSION

IMAGE:7097620 5', mRNA sequence.

## VERSION

CK365185.1 GI:40331120

## KEYWORDS

EST.

## SOURCE

Rattus norvegicus (Norway rat)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scurionathii; Muridae; Murinae; Rattus.

## REFERENCE

1 (bases 1 to 736)

## AUTHORS

NIH-MGC http://mgs.nci.nih.gov/.

## TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

## JOURNAL

Unpublished (1999)

## COMMENT

Contact: Daniela S. Gerhard, Ph.D.

## FEATURES

Office of Cancer Genomics

## SOURCE

National Cancer Institute / NIH

## FEATURES

Bldg. 31 Rm10A07 Bethesda, MD 20892

## FEATURES

Email: cgabs-remail.nih.gov

## FEATURES

Tissue Procurement: Howard Jacobs

## FEATURES

cDNA Library Preparation: Express Genomics

## FEATURES

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

## FEATURES

DNA Sequencing by: Agencourt Bioscience Corporation

## FEATURES

Clone distribution: MGC clone distribution information can be

## FEATURES

found through the I.M.A.G.E. Consortium/LNL at:

## FEATURES

http://image.lnl.gov

## FEATURES

Plate: LNL41944 Row: d Column: 02

## FEATURES

High quality sequence stop: 736.

## FEATURES

Location/Qualifiers

## FEATURES

1..736

## FEATURES

/organism="Rattus norvegicus"

## FEATURES

/mol\_type="mRNA"

## FEATURES

/db\_xref="taxon:10116"

## FEATURES

/clone="IMAGE:7097620"

## FEATURES

/tissue\_type="lung, pooled"

## FEATURES

/lab\_host="DH10B Tona"

```

/clone_lib="NIH_MGC_231"
/notes="Organ: lung; Vector: pExpress-1; Site 1: EcoRV;
Site 2: NotI; RNA obtained from pooled lung tissue from a
mix of male and female animals at 8 wk old. Tissues were
snap-frozen and kept at -80C for two days before RNA
extraction and purification (TRI-reagent method). cDNA was
primed using oligo-dT primer:
5'-pGACTGATCTTGAATCGGACGCGCCGCTT-25-3' and cloned into
the EcoRV/NotI sites of pExpress-1. Size-selection >1.4kb
resulted in an average insert size of 2.3 kb. This primary
library is not normalized (normalized primary library is
NIH_MGC 232) and was constructed by Express Genomics
(Frederick, MD). Note: this is a NIH_MGC library."
```

## ORIGIN

## Alignment Scores:

```

Pred. No.: 2.22e-76 Length: 736
Score: 749.00 Matches: 141
Percent Similarity: 91.02% Conservative: 11
Best Local Similarity: 84.43% Mismatches: 15
Query Match: 60.40% Indels: 0
DB: 7 Gaps: 0

```

US-10-649-273-2\_COPY\_176\_414 (1-239) x CK365185 (1-736)

```

QY 73 AsnGlyAspPheSerPheThrGlnGlyLeuGlnHisValThrAspLysIleIleMetLysLys 92
DB 1 AACTGATATTTCTTTTACGGGACTCAACATGTCACCGATAGCTATATACACACAG 60
QY 93 GlnLysGlnGlnGlyIleGlnLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 112
DB 61 GAAAGAAAGAGGATTTGAGAGGGGCAATCTGTCATCAGCGCGGACATTCCTGCT 120
QY 113 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 132

```

|    |     |  |     |
|----|-----|--|-----|
| Db | 121 | GGGGTACAGCAGCAACAGCGTCCACTTTGGGAAAAGAACATGTGTCTATTCTGTTT       | 180 |
| Qy | 133 | CysIysIlnArAspLeuLeuProGlnbAsnAlaValIleuValAlaSerIyGIyVal      | 152 |
| Db | 181 | TGCCAGCAGAAAATTGTGTACTCTCCAGCTAACGCACTATTGTTGTCTGGAGGTGT       | 240 |
| Qy | 153 | AlaSerAsnPhetylIleArgrAlaIeuGluIleIeuThrAsnAlaThrGlnCystr      | 172 |
| Db | 241 | GCAAGTAACTTGTACATCCGAGAGACATTGGAAATTGTAGCAAAATCAACACATCACT     | 300 |
| Qy | 173 | IeuIeuCystrProProArgrIeuCystrThrAspAsnGlyIleMetIleAlaTrpAsnGly | 192 |
| Db | 301 | TTGTTGTGTCCCCCTCCGAGACTGTGCATGCATGATCATGATTCATGAAATGGA         | 360 |
| Qy | 193 | IleGluArgrIeuArgrAlaGlyIeuGlyIleIeuHisAspIleGluGlyIleArgrIeu   | 212 |
| Db | 361 | ATTGAAAGATTACGTGCTGGCTGTGGCATTTTACAGATGATGAAACATCCCATTCGAA     | 420 |
| Qy | 213 | ProIysCystrIeuGlyValAspIleSerIySGluValGlyGluAlaSerIleIyVal     | 232 |
| Db | 421 | CCAAATATGCTCTCGGAATAGACATATCCAGAGAAATTGCAGAAAGCTGCCATPAAAGTA   | 480 |
| Qy | 233 | ProGlnIeuIyMetGluIle   | 239 |
| Db | 481 | CCACGATTAAATAGCACTT  | 501 |

Search completed: June 17, 2005, 01:30:06  
Job time : 2054.13 secs